The Construction and Use of a *Francisella tularensis* DNA Microarray

Thesis

UNRESTRICTED.

The Construction and Use of a *Francisella tularensis* DNA Microarray

A thesis submitted for the degree of Doctor of Philosophy

to the Open University

by

Helen LeButt BSc. (Hons.)

December 2007

DATE OF SUBMISSION: 13 DECEMBER 2007
DATE OF AWARD: 30 APRIL 2008.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Abstract

A DNA microarray was designed and constructed using the genome sequence of the highly virulent obligate intracellular pathogen *Francisella tularensis* strain Schu S4. The microarray was optimised and then tested by performing a comparative genomics study on *Francisella* strains. The microarray was used to distinguish between *Francisella* strains at the subspecies level, detecting differences between the genomes of the subspecies at a similar rate to differences previously published from *Francisella* comparative genomics studies. Further analysis of the genomic differences identified between subspecies using the microarray has provided some suggestions as to the genetic basis for the relative attenuation of one subspecies, and similarly, differences identified between the *F. tularensis* live vaccine strain and its progenitor strain provided some clues as the genetic basis for the attenuation of the vaccine strain. The microarray was also used to carry out functional genomics studies on *Francisella novicida* cultured under *in vitro* stress conditions: iron starvation, oxidative stress, elevated temperature, and acidic pH. A number of genes were regulated in response to each of these conditions, and a detailed analysis of the data has provided insights into the stress response of *Francisella*, and some of the mechanisms that it may employ upon encountering similar stresses *in vivo*.

Table of contents

List of figures

List of tables

List of symbols and abbreviations

A_{260}	Absorbance at 260 nm
ABI	Applied Biosystems
ACDP	Advisory Committee on Dangerous Pathogens
aCGH	Array comparative genomics
ACGM	Advisory Committee on Genetic Modification
ATCC	American Type Culture Collection
ATP	Adenosine triphosphate
bp	Base pair
BCGA	Blood cysteine glucose agar
CDS	Coding sequence
CO_2	Carbon dioxide
cm	Centimetres
CDM	Chamberlain's defined medium
CDM-Fe	Chamberlain's defined medium without iron
CDM-met	Chemically defined medium without methionine
cDNA	Complementary DNA
CFU	Colony forming units
ChIP	Chromatin immunoprecipitation
CoA	Coenzyme A
Contig.	Contiguous piece
CT	Cycle threshold
$CuCl_2$	Copper chloride
Cy3	Cyanine-3
Cy5	Cyanine-5

CyDye	Cyanine 3 and/or 5 dyes
dATP	Deoxyadenine triphosphate
dCTP	Deoxycytosine triphosphate
DEPC	Diethylpyrocarbonate
dGTP	Deoxyguanine triphosphate
DHA	Drug/hydrogen ion antiporter
DHP	Differentially hybridising probe
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTPs	Deoxyribonucleotide triphosphates
DTT	Dithiothreitol
dTTP	Deoxythymine triphosphate
ECACC	European Collection of Cell Cultures
FCS	Foetal calf serum
FCV	*Francisella*-containing vacuole
FDR	False discovery rate
Fe^{2+}	Ferrous iron
Fe^{3+}	Ferric iron
$FeSO_4$	Iron sulphate
FPI	*Francisella* pathogenicity island
FSC	*Francisella* strain collection
g	Gram
g	One unit of acceleration due to gravity on the surface of the Earth
gDNA	Genomic DNA
h	Hours

HCl	Hydrochloric acid
HPA	Health Protection Agency
HO·	Hydroxyl radical
H_2O	Water
H_2O_2	Hydrogen peroxide
HOCl	Hypochlorous acid
HSP	Heat shock protein
IFN	Interferon
inc.	Incorporated
i.p.	Intraperitoneal
IS element	Insertion element
J	Joules
KEGG	Kyoto Encyclopaedia of Genes and Genomes
l	Litre
KCl	Potassium chloride
Kg	Kilogram
K_2HPO_4	Potassium phosphate, dibasic
KH_2PO_4	Potassium phosphate, monobasic
Log	Logarithmic
LPS	Lipopolysaccharide
LVS	Live vaccine strain
M	Molar
MAPK	Mitogen-activated protein kinase
Mb	Mega bases
MFS	Major facilitator superfamily

mg	Milligram
μg	Microgram
$MgSO_4$	Magnesium sulphate
min	Minutes
ml	Millilitre
μl	Microlitre
MLD	Median lethal dose
mm	Millimetre
mM	Millimolar
μm	Micrometre
μM	Micromolar
$MnCl_2$	Manganese chloride
MOI	Multiplicity of infection
mRNA	Messenger RNA
NaAc	Sodium acetate
NaCl	Sodium chloride
NAD	Nicotinamide adenine dinucleotide
NaOH	Sodium hydroxide
NCBI	National Centre for Biotechnology information
ng	Nanogram
NLR	NACHT leucine-rich repeat
nm	Nanometre
nM	Nanomolar
NO	Nitric oxide
NOS	Nitric oxide synthase

$O_2^{\cdot -}$	Superoxide radical
OD*x*	Optical density at *x* nm
$ONOO^-$	Peroxynitrite
ORF	Open reading frame
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEP	Phosphoenolpyruvate
PMN	Polymorphonuclear cell
pmol	Picomole
psi	Pounds per square inch
PTS	Phosphotransferase
QPCR	Quantitative reverse-transcriptase PCR
RD	Region of difference
RF	Release factor
RNA	Ribonucleic acid
RNase	Ribonuclease
RND	Resistance-nodulation-cell-division
RNR	Ribonucleotide reductase
RNS	Reactive nitrogen species
ROS	Reactive oxygen species
rpm	Revolutions per minute
rRNA	Ribosomal RNA
S	Seconds
SDS	Sodium dodecyl sulphate
SOD	Superoxide dismutase

Spp.	Species
SSC	Sodium chloride-sodium citrate
Subsp.	Subspecies
TAE	Tris-acetate-EDTA
TLR	Toll-like receptor
T_m	Melting temperature
TNF	Tumour necrosis factor
UK	United Kingdom
USA	United States of America
USSR	United Soviet Socialist Republics
UV	Ultraviolet
v/v	Volume per volume
w/v	Weight per volume
∞	Infinity
>	Greater than
≥	Greater than or equal to
<	Less than
±	Plus or minus
%	Per cent
°C	Degrees Celsius
°N	Degrees North
~	Approximately

Acknowledgements

The use of microarray printers and training in these were kindly provided by FOI, and by the Health Protection Agency, Salisbury, UK. The Artemis comparison of the *F. tularensis* Schu S4 and *F. tularensis* live vaccine strain (LVS) sequences was kindly provided by Dr. M. Duffield, Dstl. Advice on microarray data analysis was kindly provided by Mr. A. Hunter, Dstl.

I would like to thank my supervisors Dr. Jo Prior, Prof. Rick Titball, and Prof. Petra Oyston for all of their guidance, help, and encouragement throughout this PhD - it has been very gratefully received.

I would like to thank my family, and in particular my parents, for their love and support, and for the interest that they have taken in the studies of their absentee daughter.

I would also like to thank my friends – you all know who you are – for their continued support and for always being there to provide a welcome distraction!

I must thank Porridge, who started this journey with me but sadly had to leave before the end, and Pudding and Pickle for their constant companionship during many hours of writing.

This thesis is dedicated to my husband Chris, for putting our first year of married life on hold and for being with me every step of the way.

Author's declaration

I declare that this thesis has been composed by myself, and that it has not been accepted in any previous application for a higher degree, that the work of which it is a record has been performed by myself unless otherwise stated and that all sources of assistance and information have been specifically acknowledged:

The preliminary annotation of the *Francisella tularensis* Schu S4 genome sequence was performed by Dr. K. Svensson, FOI, Umea, Sweden.

Oligonucleotide primers for microarray were designed by MWG Biotech Ebersberg, Germany.

Microarray analysis of the response of *Francisella novicida* to culture under iron starvation conditions was performed in collaboration with T. Milne at Dstl. However the treatment and analysis of the iron-response microarray data presented here (chapter 5a) was carried out independently and is unique to this thesis.

Publications

Work from this thesis has previously been presented or published as follows:

H. Diaper, P. Wikström, K. Svensson, M. Forsman, R.W. Titball, P.C.F. Oyston. Comparative genomics of *Francisella tularensis* using a DNA microarray. Presented at the Fifth International Conference on Tularemia, 1-4 November 2006, Woods Hole, Ma, USA and at the 160th Meeting of the Society for General Microbiology, 26-29 March 2007, Manchester, UK.

Milne, T.S., Michell, S.L., Diaper, H., Wikström. P., Svensson, K., Oyston, P.C., and Titball, R.W. 2007. A 55 kDa hypothetical membrane protein is an iron-regulated virulence factor of *Francisella tularensis* subsp. *novicida* U112. J. Med. Micro. 56 (pt 10):1268-76.

Chapter 1

Introduction

1.1. *Francisella tularensis* and tularemia

1.1.1. *Francisella*

1.1.1.1. Historical and current taxonomy

The etiological agent of tularemia (also known as rabbit- and deer-fly fever) was first isolated in 1911 in Tulare County, California following an outbreak of a plague-like illness in ground squirrels (McCoy and Chapin, 1912). The small, Gram-negative bacterium was originally placed in the genus *Bacterium* to be named *Bacterium tularense* after the county where it was isolated (McCoy and Chapin 1912), but was later placed in the genus *Pasturella*, and then subsequently in the genus *Brucella* (Wilson and Miles, 1964; Owen, 1974). Finally, in 1947 the proposal was made for the new genus of *Francisella*, with the single species *Francisella tularensis* (Dorofe'ev, 1947). The genus *Francisella* is the only member of the family *Francisellacea*, which is a member of the Thiotrichales order of the γ-proteobacteria class of bacteria.

In 1951 a similar bacterium was isolated from a water course in Utah, and in 1955 this isolate was classified as *Pasteurella novicida* (Larson *et al.*, 1955). It was not until 1959 that the isolate was re-classified as a second species of the genus *Francisella* (Olsufiev *et al.*, 1959). Based on overall identity with *F. tularensis* at the deoxyribonucleic acid (DNA) level, and 99.6% identity at the 16s ribosomal ribonucleic acid (rRNA) level, several researchers have called for the re-classification of *Francisella novicida* as a subspecies (subsp.) of *F. tularensis* (Hollis *et al.*, 1989; Forsman *et al.*, 1994). However, at the time of writing, the List of Prokaryotic Names with

Standing in Nomenclature (which draws the most up-to-date information from the Approved Lists of Bacterial Names, the International Journal of Systematic Biology, and the International Journal of Systematic and Evolutionary Microbiology) describes *F. novicida* as a separate species of the genus *Francisella* (Euzéby, 1997). Therefore, although several publications refer to *F. tularensis* subsp. *novicida*, at the time of writing this is not strictly correct, although it is fair to say that this would seem to be a species *incertae sedis*.

In 1959 another *Francisella*-like organism was isolated from a dying muskrat found in marshland, again in Utah. This isolate was 24% similar at the DNA level and was morphologically similar to *Yersinia pestis*, and so was originally classed as *Yersinia philomiragia* (Ritter and Gerloff, 1966). However serological cross-reactivity with organisms of the genera *Brucella* and *Francisella* lead to the re-classification of this organism as *Francisella philomiragia* (Ohara *et al.*, 1974; Hollis *et al.*, 1989). More recently two subsp. of *F. philomiragia* have been accepted: subsp. *noatunensis* (also known as *Francisella piscicida* sp. nov. [Ottem *et al.*, 2007]), which has so far only been isolated from fish, and subsp. *philomiragia* (Mikalsen *et al.*, 2007).

At the time of writing there are three subsp. of *Francisella tularensis*: *tularensis* (sometimes known as *nearctica* or 'type A'), *holarctica* (sometimes known as *palaearctica* or 'type B'), and *mediaasiatica* (fig. 1.1). The three subsp. are broadly separated by geographical location and by altered virulence in man (see section 1.1.2.). Subsp. *holarctica* can be further

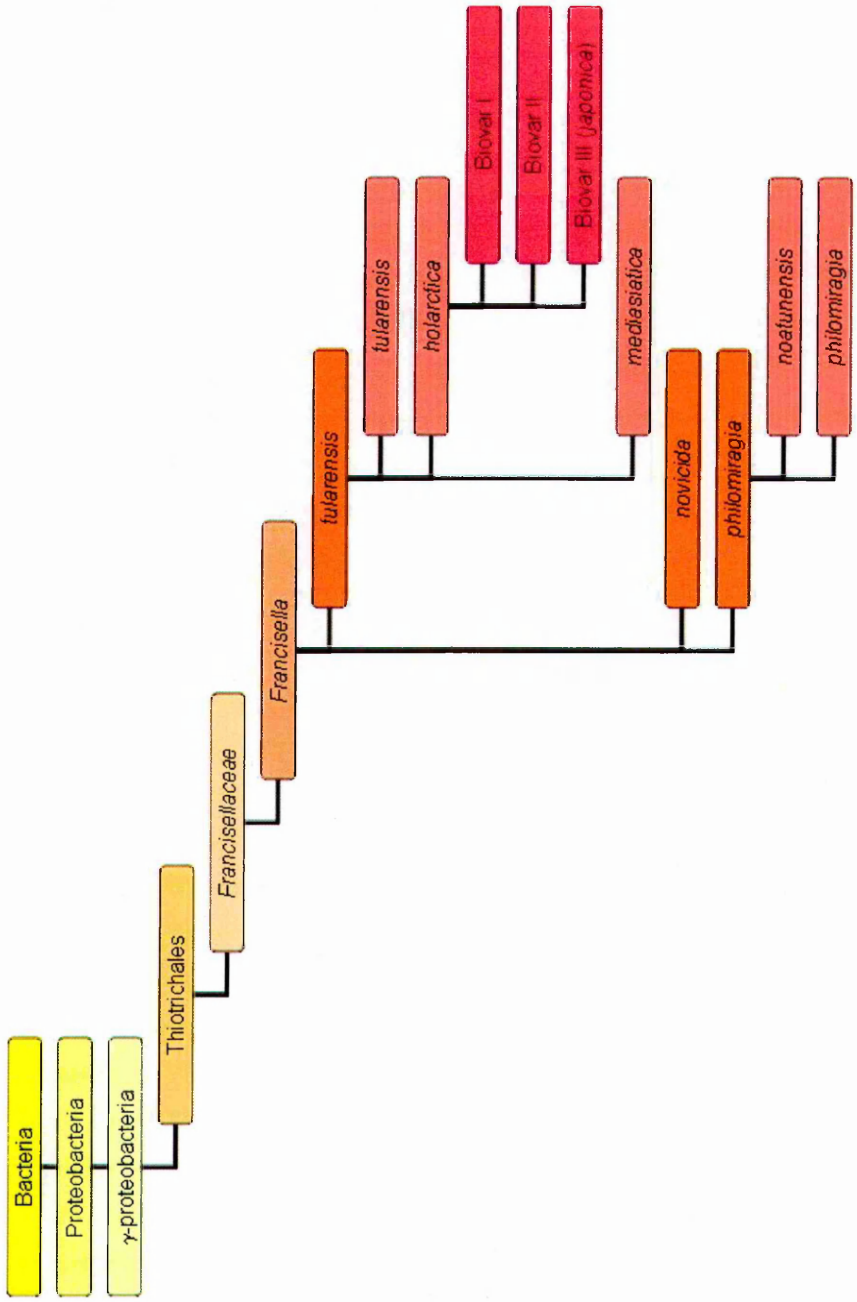


Fig. 1.1 The taxonomy of *Francisella*

Kingdom	Phylum	Class	Order	Family	Genus	Species	Subsp.	Biovar

divided into three biovars: I, which is erythromycin sensitive; II, which is resistant to erythromycin, and III, isolates from Japan (also known as biovar *japonica*), and subsp. *tularensis* can be further divided into two distinct clades, A.I. and A.II., which are based on variable-number tandem repeat analysis and broadly on geographical location (Johansson *et al.*, 2004).

1.1.1.2. Genome sequence

The first whole genome sequence of a highly virulent *F. tularensis* subsp. *tularensis* strain to be completed was that of strain Schu S4 (*Francisella* strain collection [FSC] 237) (Larsson *et al.*, 2005). The genome size was reported as 1.89 mega bases (Mb), and was predicted to contain 1852 gene coding sequences (CDS), of which 302 have no previously reported homologues. Few classical virulence factors are encoded apart from a putative polysaccharide capsule, and lipopolysaccharide (LPS), however the relatively large number of *Francisella*-specific CDS may encode novel virulence factors. Of the 1852 CDS, approximately 10% are pseudogenes (14% of which were disrupted by insertion [IS] elements). Some 50% of the predicted metabolic pathways are disrupted, which indicates that *F. tularensis* is undergoing genome reduction, and may explain why 14 separate compounds are required to culture *Francisella* in the laboratory (Chamberlain, 1965). At the time of writing, the genomes of a further seven strains of *Francisella*, including *F. novicida*, and *F. tularensis* subsp. *tularensis* and subsp. *holarctica* have been sequenced, and a further seven sequencing projects are in progress.

1.1.1.3. Geographical distribution

All species of *Francisella* are widely distributed in the Northern hemisphere, traditionally only between latitudes 30°N and 70°N (Hopla, 1974; Forsman *et al.*, 1990) (fig. 1.2), however the isolation of a *F. novicida*-like organism from the Northern Territory of Australia has cast some doubts to this dogma (Whipp *et al.*, 2003).

All isolates of *F. novicida*, with the exception of the *F. novicida*-like organism from Australia (Whipp *et al.*, 2003) have originated in North America. Likewise, all isolates of *F. philomiragia* subsp. *philomiragia* have come from North America, although *F. philomiragia* subsp. *noatunensis* has only been isolated from cod in the North Atlantic Ocean (Mikalsen *et al.*, 2007; Ottem *et al.*, 2007).

North America is also the main location where *F. tularensis* subsp. *tularensis* is found, with the only isolates of this subsp. in Europe almost certainly attributable to environmental contamination with a commonly used laboratory strain (Chaudhuri *et al.*, 2007). Clade A.I. of *F. tularensis* subsp. *tularensis* is found in the Midwest, California, and Massachusetts, whilst clade A.II. is found in California and the mountain states (Johanssen *et al.*, 2004). The most widespread subsp. of *F. tularensis* is *holarctica*, which is the predominant subsp. in Europe, but is also found in North America and Asia. Of the three biovars of *F. tularensis* subsp. *holarctica*, I and II have been isolated in North America and Eurasia, whereas biovar *japonica* is confined to Japan. There have been recent publications of emerging or re-

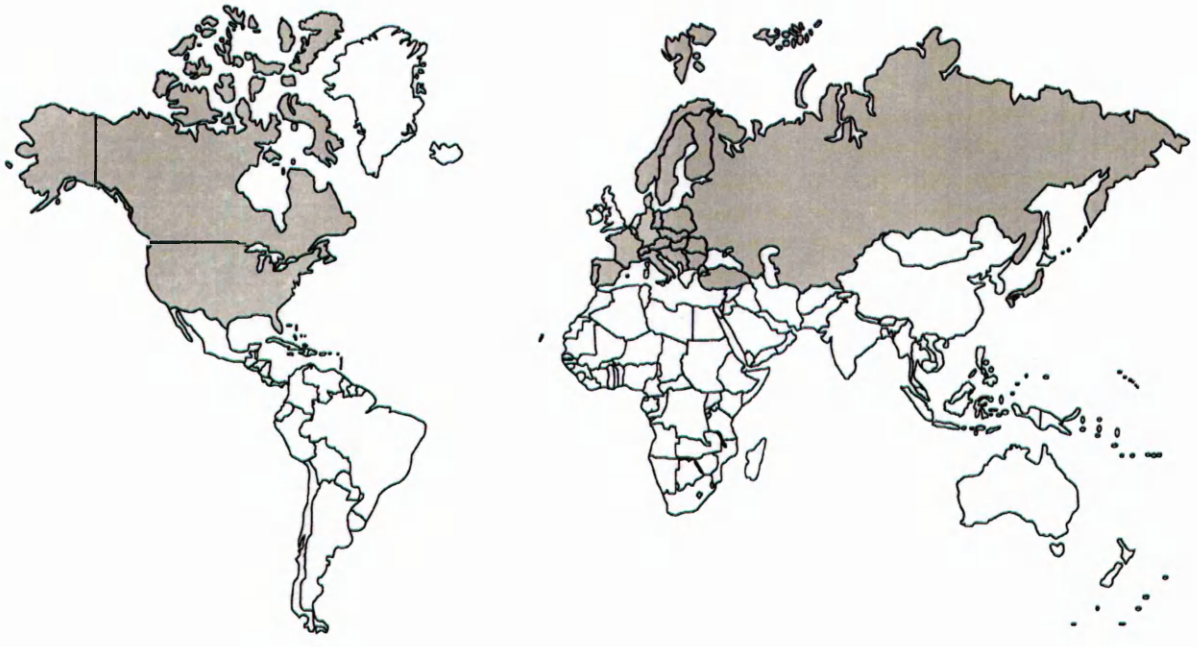


Fig. 1.2 World map showing *F. tularensis*-endemic regions, adapted from Ellis *et al.*, (2002). Both *F. tularensis* subsp. *tularensis* and subsp. *holarctica* are endemic to North America. Only *F. tularensis* subsp. *holarctica* is endemic to Eurasia. Both *F. tularensis* subsp. *holarctica* and subsp. *mediaasiatica* are endemic to the central Asian republics of the former USSR. There is also a report of *Francisella tularensis* subsp. *holarctica* isolated from China (Zhang *et al.*, 2006), and a report of a *Francisella novicida*-like organism isolated from Australia (Whipp *et al.*, 2003).

emerging *F. tularensis* subsp. *holarctica* in countries including Germany, Denmark, and China, which would seem to indicate that it is more widespread than previously thought (Byström *et al.*, 2005; Splettstoesser *et al.*, 2007; Zhang *et al.*, 2006). The third *F. tularensis* subsp., *mediasiatica*, is only found in the central Asian republics of the former United Soviet Socialist Republics.

1.1.2. Tularemia

1.1.2.1. Disease features in man

Disease in previously healthy humans is usually caused by *F. tularensis* subsp. *tularensis* or *F. tularensis* subsp. *holarctica*, with the former being of high virulence and the latter of moderate virulence. Infection with *F. tularensis* subsp. *mediasiatica* only causes mild disease in humans (Hubálek *et al.*, 2004).

Rarely, human disease can be caused by *F. philomiragia* and *F. novicida*. Only three cases of human infection with *F. novicida* have been reported including the Australian *F. novicida*-like isolate mentioned in section 1.1.1.3. (Hollis *et al.*, 1989; Whipp *et al.*, 2003). Human disease caused by *F. philomiragia* has only been reported in patients from three risk groups: chronic granulomatous disease, myeloproliferative disease, and victims of near-drowning (Wenger *et al.*, 1989).

The overall rate of mortality from untreated infections with *F. tularensis* subsp. *tularensis* ranges from 5% to 15% (Dennis *et al.*, 2001), although the

mortality rate varies according to the dose and site of inoculum, for example, the most severe (untreated) systemic and pneumonic forms of the disease give rise to between 30% and 60% mortality (Saslaw *et al.*, 1961). Infection with *F. tularensis* subsp. *holarctica* seldom results in death in humans, with a mortality rate of less than (<) 5% (Dennis *et al.*, 2001; Ellis *et al.*, 2002).

1.1.2.1.1. Pneumonic tularemia

Primary pneumonic tularemia, resulting from direct inhalation of contaminated aerosols, is the most severe form of the disease. Inhalation of less than (<) 10 colony forming units (CFU) of *F. tularensis* is sufficient to cause disease in humans (Saslaw *et al.*, 1961). Pneumonic tularemia may also occur by secondary haematogenous spread from a distal site of infection. Symptoms may include one or more of the following: pharyngitis, bronchiolitis, pleuropneumonitis, and hilar lymphadenitis. Alternatively clinical signs may be manifest as those of systemic illness without overt signs of respiratory disease.

1.1.2.1.2. Ulceroglandular tularemia

Ulceroglandular tularemia is the most common form of the disease, but is generally less serious than the pneumonic form, with a mortality rate of < 3% (Evans *et al.*, 1985). After three to six days incubation following infection (usually a bite from an infected arthropod or transmission via a break in the skin from handling infected materials), a tender cutaneous ulcer appears at the site of inoculation. At about the same time there is onset of a febrile illness which may be characterised by chills, fever, head- and generalised body-aches (Evans *et al.*, 1985; Ohara *et al.*, 1991). Bacteria

disseminate from the site of infection via the lymphatic system to the regional lymph nodes, from where they are further disseminated to distil organs such as the liver, spleen, kidneys, and lungs. Typically lymphadenopathy results, with one or more of the regional lymph nodes becoming enlarged and tender. Glandular (as opposed to ulceroglandular) tularemia is characterised by lymphadenopathy but without the presence of an ulcer.

1.1.2.1.2.1. Oculoglandular tularemia

In this rarely seen variation of ulceroglandular tularemia, the conjunctiva is the initial site of infection, with bacteria thought to be transferred to the eyes via the fingertips (Steinemann *et al.*, 1999). Ulceration occurs on the conjunctiva, accompanied by pronounced chemosis, vasculitis, and regional lymphadenitis (Steinemann *et al.*, 1999).

1.1.2.1.3. Oropharyngeal and gastrointestinal tularemia

Ingestion of infected foodstuffs or liquids can lead, depending on the site of colonization, to either oropharyngeal or gastrointestinal tularemia. Patients with oropharyngeal tularemia usually develop a painful sore throat due to exudative pharyngitis or tonsillitis, sometimes with ulceration (Reintjes *et al.*, 2002). As with the other disease presentations, bacteria are trafficked to the regional lymph nodes resulting in this case in either cervical or retropharyngeal lymphadenopathy (Reintjes *et al.*, 2002). Gastrointestinal tularemia often presents with anorexia, nausea, vomiting, and diarrhoea. The diarrhoea is rarely bloody but may be accompanied by ulceration of the intestine. In prolonged cases enlargement of the liver and spleen may also

occur, and hepatic involvement occurs in up to 75% of patients (Zaidi and Singer, 2002).

1.1.2.1.4. Typhoidal tularemia

Typhoidal tularemia is the term used to describe acute systemic illness with no obvious clues as to site of inoculation or anatomical localisation of infection. Therefore unlike the other disease presentations, typhoidal tularemia can be characterised by a lack of lymphadenopathy (Evans *et al.*, 1985).

1.1.2.1.5. Sepsis

Sepsis resulting from tularemia is potentially fatal, and may initially present like typhoidal tularemia as a non-specific febrile illness. Patients may display acute symptoms of sepsis, including fever, chills, tachycardia, and tachypnea, they may also appear confused or comatose. Without prompt treatment patients may develop septic shock leading to acute respiratory distress syndrome and eventual organ failure (Sunderrajan *et al.*, 1985).

1.1.2.2. Vaccination and treatment

A live attenuated vaccine strain of *F. tularensis* subsp. *holarctica* (LVS), also denoted FSC155 has been shown to protect human volunteers against inhalational challenge with ten infectious doses of *F. tularensis* subsp. *tularensis*, although only partial protection was afforded against 100 and 1000 infectious doses (McCrumb, 1961). It is thought that LVS was itself derived from a vaccine strain, known as 'strain 15 restored' (currently denoted FSC338) which was attenuated by continuous *in vitro* subculture at

the Gamaleia Institute in the USSR in the 1930's where it was used to immunise tens of thousands of people living in tularemia endemic areas (Tigertt, 1962). In 1956 vaccine strains including 'strain 15 restored' were transferred from the Gamaleia Institute to the United States Army Medical Research Institute of Infectious Diseases (Tigertt, 1962), where 'strain 15 restored' was subjected to serial animal passage for the isolation of a culture that was ultimately selected for vaccine production in the United States of America (USA) and introduced as *F. tularensis* LVS (Eigelsbach and Downs, 1961). LVS was used, under the status of Investigational New Drug, to immunise laboratory workers, with a subsequent reduction of the incidence of laboratory acquired tularemia at the United States Army Medical Research Institute of Infectious Diseases from 5.7 cases per 1000 person years of risk to 0.27 cases per 1000 person years of risk (Burke, 1977). However, it has been reported that some individuals vaccinated with LVS go on to develop tularemia, and also that the immunogenicity of LVS can be variable due to sub-populations of LPS O-antigen-deficient bacteria present in vaccine cultures (Eigelsbach and Downs, 1961; Hartley *et al.*, 2006). The molecular basis of the attenuation of *F. tularensis* LVS remains unclear. The genome sequence of LVS (Chain *et al.*, 2006) has revealed > 90% identity to that of the fully virulent *F. tularensis* subsp. *tularensis* strain Schu S4, although by comparison with that of strain Schu S4, the genome of LVS appears to have undergone significant rearrangement. One genomic region that is missing in LVS compared to Schu S4 is associated with the production of type IV pili, which may partly account for the attenuation of the vaccine strain, particularly in light that a *F. tularensis* subsp. *holarctica* Δ*pilA* strain was also shown to

be attenuated (Forslund *et al.*, 2006). However, with the issues of safety, efficacy, and the undefined nature of the vaccine in mind, the US Food and Drug Administration have rescinded the Investigational New Drug status of LVS, and, at the time of writing, an alternative vaccine is sought.

The recommended treatment for confirmed cases of tularemia is with an aminoglycoside as relapses or primary treatment failures using this class of antibiotics are relatively fewer than for treatment with bacteriostatic agents (Dennis *et al.*, 2001). Previously either streptomycin or gentamicin were recommended for treatment of tularemia (Dennis *et al.*, 2001), however due to adverse side effects associated with the use of streptomycin, gentamicin has become the antibiotic of choice (Hassoun *et al.*, 2006). Bacteriostatic agents are recommended for occasions when *F. tularensis* infection is suspected but not confirmed and for mass casualty situations, with doxycycline, ciprofloxacin, or chloramphenicol being the preferred choices (Dennis *et a*l., 2001).

1.1.2.3. Epidemiology

Tularemia has been isolated from more than 250 animal species, including mammals, birds, fish, amphibians, arthropods and protozoa (Mörner, 1992). Mammalian hosts of *Francisella* are very diverse: lagomorphs including rabbits and hares, mustelids including skunks, and rodents, including voles, mice, water rats, squirrels, muskrats, beavers, gerbils, and lemmings. Antibodies to *F. tularensis* have also been found in mule deer and American elk (wapiti) (Merrell and Wright, 1978), and in

coyotes and raccoons (Bischof and Rogers, 2005). There have also been reports of infection acquired from companion animals such as hamsters, cats and prairie-dogs (Avashia *et al.*, 2004; Magnarelli *et al.*, 2007). Mammals, including humans, usually acquire the disease via arthropod transmission (ticks, flies, mosquitoes), although infection can also occur via direct contact with contaminated environments including water, soil and vegetation (Gordon *et al.*, 1983). It is thought that survival of *Francisella* in water courses may be facilitated by a symbiotic relationship with amoebae, for example *Francisella* is able to multiply in *Acanthamoeba castellanii* (Abd *et al.*, 2003). No person-to-person transmission of tularemia has been reported.

Tularemia most often affects individuals involved in rural occupations such as hunting, trapping, and farming, although laboratory workers are another group at higher than normal risk as they may be susceptible to laboratory acquired infection (Martin *et al.*, 1982; Rusnak *et al.*, 2004). Human outbreaks of tularemia may be preceded by enzootic cycles and extensive host die-offs (Zaidi and Singer, 2002). In the USA outbreaks of tularemia in man follow a bimodal distribution, with peaks occurring in the summer which are associated with ticks and in the winter which are associated with animal contact (Zaidi and Singer, 2002).

The ability of *Francisella* to survive in water courses may have contributed to major human outbreaks of the disease in Scandinavia and in Russia (Berdal *et al.*, 1996; Rogutskiĭ *et al.*, 1997). The largest naturally occurring outbreak of airborne disease to date occurred in Sweden in 1966-7,

where more than 600 farmers became infected with *F. tularensis* subsp. *holarctica,* mainly through inhalation of contaminated aerosols (Dahlstrand *et al.*, 1971). There is some evidence to suggest that the annual average number of cases of tularemia may be increasing, for example the average number of cases per year in Southern Sweden from 1999 to 2005 was 22, however this figure increased to 90 reported cases by the end of the first eight months of 2006 (Wik, 2006).

1.1.2.3.1. Illegitimate use

It is because of its low infectious dose, extreme virulence with potentially high mortality and relative ease of dissemination that *F. tularensis* is included on the (USA) Centres for Disease Control and Prevention Category A list of agents, to be considered a potential biological weapon. A World Health Organization expert committee commissioned to assess the impact of and emergency response to the illegitimate use of *F. tularensis* as a weapon estimated that aerosol dispersal of 50 kg over an urban population of five million people would result in a quarter of a million casualties and 19 thousand deaths (Dennis *et al.*, 2001).

The 1925 Geneva protocol prohibits the use of biological weapons, as does the Biological and Toxin Weapons Convention, which opened for signatures in 1972 and was implemented in 1975. However, it is known that *F. tularensis* has previously been studied under a Japanese weapons programme (Harris, 1992). There is also some evidence to suggest that *F. tularensis* may have been deliberately used by the USSR against the Germans at the battle of Stalingrad (Alibeck, 1999), although given the

unsanitary conditions of battle a naturally occurring epidemic cannot be ruled out (Croddy and Krcalova, 2001). During the Cold War both the USA and the USSR stockpiled tons of infectious agents, including *F. tularensis* (Christopher *et al.*, 1997). In advance of the Biological and Toxin Weapons Convention, the USA terminated its biological weapons programme in 1970, and destroyed its stockpiles by 1973 (Christopher *et al.*, 1997). The USSR broke up in 1991, but the fate of remaining stocks of infectious agents is not fully documented. As of June 2005, 171 sovereign states were signatories to the Biological and Toxin Weapons Convention, of which 155 had ratified it (www.opbw.org/). However, 23 sovereign states remained non-signatories, and this combined with a continued threat from non-government-sponsored terrorist organisations means that the possibility of *F. tularensis* being used as a biological weapon cannot be discounted completely.

1.1.3. The intracellular lifestyle of *Francisella*

1.1.3.1 Most invading bacteria are detected and destroyed by phagocytes

The word phagocytosis comes from the Greek *'phagein'*, meaning 'to eat' and monocytes and macrophages, which are the first-line guards of the innate host-defence system, do precisely that as they recognize and phagocytose invading micro-organisms. Cells of the innate immune system use toll-like receptors (TLRs) to sense bacterial invasion through recognition of molecular patterns. Detection of a pathogen induces the release of many cytokines (pro- and anti-inflammatory which together provide a balanced inflammatory response) which summon further phagocytes to the site of infection. Once internalized, bacteria are contained within a membrane-

bound vesicle known as the phagosome, which eventually fuses with a lysosome to form a phagolysosome where bacteria are exposed to a variety of conditions designed to kill them.

The antimicrobial action of phagocytes can be classed as either oxygen-dependent or oxygen-independent: the oxygen-dependent mechanism relies upon the generation of reactive oxygen species (ROS) within the phagosome. The oxygen and adenosine triphosphate (ATP) required for this so-called oxidative burst are produced during the process of phagocytosis (Carnutte and Barbior, 1974). The oxygen-independent mechanism results in acidification of the phagosome (to pH 5.5) and causes the release of hydrolytic enzymes and defensins (small cationic peptides which permeabilize cell membranes) (Kagan *et al.*, 1990). Other adverse intracellular conditions include nutrient and iron deprivation.

Neutrophils (a class of polymorphonuclear cell [PMN]) are short-lived phagocytic cells containing cytoplasm-located defence molecules and are highly efficient killers of bacteria. They rely largely on opsonins for phagocytosis and bactericidal activity is strictly oxygen-dependent, as these cells generate large quantities of ROS (including hypochlorous acid [HOCl]), and hypophalite ions (chloride, iodide and bromide).

1.1.3.2. Some bacteria can resist killing by phagocytic cells

The ability to survive in what would normally be considered a stressful environment for micro-organisms is an important factor in virulence of

intracellular pathogens. Bacteria that are adapted to the intracellular environment must regulate their gene expression in response to nutrient availability, pH, osmolarity, growth phase, oxygen tension, iron levels, and temperature which together constitute a global stress condition (Mekalanos, 1992). In general, bacteria have three ways of surviving inside macrophages: 1) via the extraphagosomal pathway whereby bacteria escape from the phagosome before phagosome-lysosome fusion, for example *Listeria monocytogenes* (de Chastellier and Berche, 1994); 2) via the phagolysosomal pathway whereby bacteria have become adapted for survival inside the phagolysosome (or at least inside vacuoles possessing some lysosomal markers), for example *Salmonella enterica* serovar Typhi (Ishibashi and Arai, 1995); and 3) via the intraphagosomal pathway whereby bacteria are able to block phagosome-lysosome fusion, for example *Mycobacterium tuberculosis (*Clemens and Horwitz, 1995).

1.1.3.3. Infection by *F. tularensis*

Clearly, in order to survive in the intracellular environment *F. tularensis* must subvert the normal function of the host phagocytic cell and in turn be provided with protection from the immune system. The broad mechanism of survival of *Francisella* in macrophages is reported as via the intraphagosomal pathway, with subsequent escape from the phagosome (Anthony *et al.*, 1991; Clemens *et al.*, 2004). As yet the precise mechanisms by which *Francisella* avoids being killed under conditions that for many bacterial species would prove fatal remain largely un-elucidated, but that this obligate intracellular pathogen is highly adapted for intracellular survival cannot be in question.

The variety of mammals that *F. tularensis* is capable of infecting includes human and non-human primates, rat, rabbit, guinea pig and mouse (Sjöstedt, 2006). There is also some evidence that *F. tularensis* can replicate in non-professional phagocytes, for example fibroblasts and hepatocytes tick epithelial cells, endothelial cells and HeLa cells (Tarnvik 1989; Conlan 1992).

1.1.3.3.1. Uptake of *F. tularensis*

In general the O-antigen portion of the LPS of Gram-negative bacteria is thought to provide protection from complement-mediated killing by serum. LPS also induces a proinflammatory immune response through interaction of the lipid A portion of LPS with Toll-like receptor 4 (TLR4). However, *F. novicida* LPS, which shares an identical lipid A portion with all three *F. tularensis* subsp., is not recognised by human or murine TLR4, and does not stimulate TLR4-mediated responses in mice (Hajjar *et al.*, 2006). The *F. tularensis* LPS has been shown to activate complement (Fulop *et al.*, 1993), and complement and C36 complement receptor have been shown to be a requirement for the internalization of *F. tularensis* (Löfgren *et al.*, 1983; Clemens *et al.*, 2005). Taken together this means that *Francisella* LPS could be described as benign but necessary for infection.

Uptake of *F. tularensis* by macrophages requires the presence of opsonins (Allen, 2003). It has been demonstrated that macrophage uptake of both live and dead *Francisellae* is via a novel form of pseudopod looping in

such a way that bacteria initially reside inside a large vacuole which rapidly shrinks to the more tightly fitting phagosome (Clemens *et al.*, 2005).

Ingestion of *F. tularensis* by PMNs also occurs in the presence of serum opsonins (Allen, 2003), however there is conflicting published evidence as to the extent of survival of *F. tularensis* that has been ingested by neutrophils: McCaffrey and Allen report that *F. tularensis* LVS ingested by PMNs is not eliminated, and goes on to escape the phagosome and replicate in the cytoplasm. It is hypothesised by these authors that this occurs because *F. tularensis* is able to disrupt the respiratory burst (McCaffrey and Allen, 2006). This may be possible because *F. tularensis* possesses AcpA, a respiratory burst-inhibiting acid phosphatase, which has been shown to be active in PMNs (Reilly *et al.*, 1996). However AcpA was found to be non-essential for survival of *F. novicida* in macrophages (Baron *et al.*, 1999), although this may represent a difference between the mechanisms by which *F. tularensis* and *F. novicida* survive inside PMNs. It has also been suggested that further disruption of the respiratory burst may be because *F. tularensis* LPS is not recognized by the LPS-sensing molecules of PMNs (Barker *et al.* 2006).

In contrast to the above reports, it has previously been reported that neutropenic mice are susceptible to an otherwise sub-lethal dose of *F. tularensis* LVS (Sjöstedt *et al.*, 1994) and that a neutropenic human bone marrow transplant patient acquired a fatal *F. tularensis* infection (Sarria *et al.*, 2003). Together these two reports would seem to indicate that killing by

PMNs is essential for controlling *F. tularensis* infection in mice and humans. Furthermore, it has been reported that a majority (98%) of *F. tularensis* LVS are killed by PMNs, whereas 63% of fully virulent *F. tularensis* subsp. *holarctica* survive (Löfgren *et al.* 1983). The same group has shown that, *in vitro*, *F. tularensis* LVS is much more sensitive to HOCl than fully virulent *F. tularensis* subsp. *holarctica*, which may account for the increased susceptibility of the vaccine strain to killing by PMNs (Löfgren *et al.*, 1984).

1.1.3.3.2. Inside the phagosome

After uptake by macrophages, *F. tularensis* resides within a phagosome. It appears that *F. tularensis* is able to alter the maturation of the phagosome (evidenced by the exclusion of cathepsin D, lysosomal tracers and lack of acidification) and subsequently prevent phagolysosome fusion (Clemens *et al.*, 2004). Within two to four hours of invasion bacteria begin to escape the phagosome by exporting surface-located material in vesicles which then associate with the phagosomal membrane and cause it to degrade (Golovliov *et al.*, 2003; Lindgren *et al.*, 2004).

1.1.3.3.3. Outside the phagosome

Within four to eight hours of invasion most *F. tularensis* have escaped the phagosome to reside in the cytoplasm, with access to the intracellular milieu for growth and replication. Replication of bacteria begins six to 12 hours after infection and the resulting high levels of viable bacteria induce cytopathogenesis and apoptosis via a pathway that partly resembles the intrinsic apoptotic pathway: *F. tularensis* LVS has been shown to induce the

release of cytochrome C from the mitochondria of murine J774A.1 macrophages, changing the potential of the mitochondrial membrane (Lai and Sjöstedt, 2003). Activation of Caspase 9 and Caspase 3 has also been observed in *F. tularensis* LVS-infected J774A.1 cells (Lai and Sjöstedt, 2003). Apoptosis of macrophages containing *F. tularensis* LVS requires activation of the p42/p44 mitogen-activated protein kinase (MAPK) signal pathway, which is known to be stimulated by LPS. Apoptosis of macrophages containing *F. tularensis* LVS is also associated with reduced activity of the p38 MAPK pathway (Hrstka *et al.*, 2005). The *F. tularensis* LVS gene *iglC* has been shown to be required for apoptosis to occur (Lai *et al.*, 2004) and a Δ*purMCD* mutant of the same strain has been shown to escape from the phagosome but not to replicate in the cytoplasm or to induce apoptosis (Pechous *et al.*, 2006).

The results of a recent study carried out by Checroun *et al.* suggest autophagy as an alternative means by which *F. tularensis* may escape from host cells (Checroun *et al.*, 2006). The study, carried out using *F. tularensis* subsp. *holarctica*-infected murine primary macrophages, showed the formation of double-membrane bound *Francisella*-containing vacuoles (FCVs) in response to cytoplasmic replication of *F. tularensis*. FCVs were shown to be autophagic in nature, and it was demonstrated that fusion with lysosomes and acidification of FCVs occurred, although bacteria contained within FCVs were morphologically intact (Checroun *et al.*, 2006). It is not explicitly clear from this study whether or not the formation of autophagic FCVs is in fact xenophagy, mediated by the host cell in order to limit the

proliferation of bacteria in the cytoplasm, or whether this is a bacteria-mediated event that facilitates exit from the host cell through exocytosis. The visualisation of intact bacteria inside FCVs and the fact that biogenesis of FCVs is dependent upon active proliferation of *F. tularensis* are important indicators that this may be another example of how *F. tularensis* is able to subvert host cell defences. Several other intracellular bacterial species have been shown to survive the autophagic pathway (Dorn *et al.*, 2002), although, at the time of writing, *F. tularensis* is unique in its apparent escape from and subsequent re-entry into two distinct endocytic pathways (Checroun *et al.*, 2006).

A class of intracellular molecules (like TLRs), called NACHT leucine-rich repeats (NLRs), recognise and respond to internalized pathogens. Upon *F. tularensis* internalization, NLRs induce an inflammatory response mediated by interleukin 1β which leads to the recruitment of fresh macrophages for infection once apoptosis (or presumably exit via the autophagic route) has occurred at 24 to 48 hours after internalization (Gavrillin *et al.*, 2006). These macrophages are not activated and therefore easier to infect by bacteria escaping from host cells.

A recent publication has cast some doubt on the long-standing paradigm that *F. tularensis* in the blood of infected hosts is primarily cell-associated. Analysis of blood taken from mice infected with either *F. tularensis* LVS or Schu S4 showed that an average of 75% of bacteria recovered were located in the plasma rather than intracellularly (Forestal *et*

al., 2007). However this study failed to investigate whether *F. tularensis* is able to replicate extracellularly, or whether the extracellular bacteria observed were simply escapees from phagocytic cells, preparing for another round of infection as described previously.

No known exotoxins are produced by *F. tularensis*, instead the pathogen would appear to usurp the immune system and use it against the host. If death is the ultimate outcome of infection this is usually due to pneumonia and/or septic shock brought about by the release of large quantities of cytokines (Dennis *et al.*, 2001).

1.2. DNA microarrays

DNA microarrays have gained wide use in biological research as a powerful technology enabling the simultaneous analysis of a large number of nucleic acid hybridisations. DNA representing genes of interest are immobilised or arrayed, in an addressable format, onto treated glass, usually either a slide or a silicon chip. DNA arrayed onto the solid surface are conventionally referred to as probes (Shalon *et al.*, 1996). Sample DNA, either genomic (gDNA) or complementary (cDNA) depending on the experimental design, is then hybridised to the microarray under stringent conditions for the detection of sequence specific binding. Labeled DNA that is hybridised to the microarray is conventionally referred to as target DNA (Shalon *et al.*, 1996). Broadly, microarrays have two main utilities: comparative- and functional genomics.

1.2.1. Microarray formats

At the time of writing there are two main microarray formats available commercially, based on glass slides and silicon chips. Silicon GeneChip technology is owned by Affymetrix Incorporated (Inc.) (Santa Clara, USA), and consists of extremely dense oligonucleotide arrays that are synthesised *in situ* by a process known as photolithography. Affymetrix have set the gold standard in microarray chip technology - each gene being represented by 25 unique oligonucleotides and internally controlled by 25 mis-matched oligonucleotides which differ by one nucleotide each. Affymetrix GeneChip technologies, however, are among the most expensive in the field.

For microarrays printed, or spotted, onto glass slides, the probes may be cDNA directly amplified from the genome by polymerase chain reaction (PCR), or they may be oligonucleotides, usually between 25 and 80 nucleotides in length, that are designed *in silico* to be unique to each CDS. Both cDNA microarrays and oligonucleotide microarrays have been shown to perform comparably well in bacterial transcription studies (Guckenberger *et al.*, 2002), and the relatively low cost of glass slide microarrays compared to Affymetrix GeneChips makes them attractive to many researchers.

Alternative technologies to both Affymetrix GeneChips and glass slide microarrays are becoming available; one example of this is BeadChip technology from Illumina, Inc. (San Diego, USA): clusters of 50,000 beads are arrayed into micro-wells on glass slides, with each bead containing more than 10^5 copies of a covalently bound 50-mer oligonucleotide probes. Each

slide contains several arrays, with each array containing more than 1500 unique oligonucleotide sequences. Very high-density microarrays of this type allow high-throughput genetic profiling, with the capability to include multiple organisms on one chip.

1.2.2. A glass-slide microarray experiment using fluorescent labeling

A microarray experiment can be broken down into four key stages, at each of which great care must be taken to avoid the introduction of systematic errors:

1.2.2.1. Fabrication of microarrays

Using a sequenced, and ideally, annotated genome, unique probes are designed corresponding to each CDS of interest. When considering the length of probes, there are two main factors to be taken into account: sensitivity, and specificity. Longer probes are likely to yield higher sensitivity but with reduced specificity, whilst for shorter probes the opposite is true. The probes are checked *in silico* for likelihood of cross-hybridisation. There are two commonly accepted criteria to indicate that the probability of cross-hybridisation is low: < 75% overall complementarity between probe and 'non-target' gene sequences; and <15 contiguous complementary bases between the two (Kane *et al.*, 2000).

A robot, facilitated by microelectronic and miniaturisation technologies is used to array each probe onto a chemically treated glass slide. There are several chemical treatments available for glass slides, the most appropriate

being dependent on probe type, some of these are discussed in more detail in section 3.4.2. Where necessary because of the surface chemistry of the microarray, probes are then further immobilised by cross-linking, either by ultraviolet (UV) treatment, or by baking to form a covalent bond.

1.2.2.2. Sample preparation and labeling

Target nucleic acid from the organism of interest is isolated by standard DNA or ribonucleic acid (RNA) purification methods. Care must be taken at this stage, particularly in the case of RNA, which has a very short half-life, not to introduce artefacts through the process of transcription or ribonuclease (RNase) activity. DNA is labeled using fluorescent nucleotides via random prime incorporation during a polymerase reaction. RNA is reverse-transcribed into cDNA, and again the labeled nucleotides are incorporated by random priming. The fluorescent dyes Cy3 and Cy5 are often used in microarray experiments, particularly as it is possible to purchase nucleotides labeled with these dyes from GE Biosciences (Amersham, UK). In addition to fluorescence, microarray readout can also take the form of radioisotope detection and mass spectrometry (Cox *et al.*, 2001; Isolar *et al.*, 2001).

1.2.2.3. Hybridisation, stringency washing and imaging

The fluorescent-labeled target DNA is hybridised under stringent conditions to the arrayed probes before washing to remove unbound DNA. Both hybridisation and stringency washing conditions are determined experimentally, and are strongly influenced by the melting temperature of the

probes, as in all nucleic acid hybridisation reactions. After stringent washing, the microarray is scanned using lasers at the correct wavelength to excite the fluorophore(s) and a digital image is recorded. By differentially labeling nucleic acid samples with fluorophores requiring different excitation wavelengths, it is possible to competitively hybridise two target DNA species to the microarray. The excitation maximum of Cy3 is at 565 nm and the excitation maximum of Cy5 is at 670 nm. The digital image of the scanned microarray is pixelated, and this information is used to determine the relative hybridisation of different target samples.

1.2.2.4. Normalisation and statistical analysis

There are various methods for mining the data generated from hybridised microarrays, ranging from very simple qualitative (spot is fluorescing or not) to more complex statistical analyses. One important consideration to be made before statistical analysis can proceed is the correction of dye-dependent bias in signal - Cy3 has a stronger signal than Cy5, causing artefactual variation in signal. One method to eliminate this is to perform 'dye-swap' experiments, where each target is separately labeled with both Cy3 and Cy5 and both samples are hybridised with their differentially labeled counterpart. The resulting scan data is pooled, essentially smoothing the signal. Alternatively, this smoothing can be achieved by performing other normalisation calculations, for example by performing a lowess transformation (Yang *et al.*, 2002). The correct statistical analysis of an experiment is dependent on the experimental design, and should be taken into account during the preliminary stages of

any study. Although microarray analysis is a quantitative procedure, the generally accepted method for confirming up- or down regulation of the transcription of individual genes is by quantitative-reverse transcriptase-PCR (QPCR).

1.2.3. Comparative genomics

1.2.3.1. Whole-genome sequencing

At the point of writing the National Centre for Biotechnology Information (NCBI) database holds the complete annotated sequences of more than 300 bacterial genomes (www.ncbi.nih.gov). The first bacterial genome to be sequenced was that of *Haemophilus influenza* (Fleischmann *et al.*, 1995) with early sequencing projects reported to have taken approximately 18 months. Current sequencing technologies have dramatically reduced the time taken to complete a genome sequence, with a 3.0 Mb sequence taking approximately 4 months to complete, including annotation. The relative speed and ease of completing a bacterial genome sequence make this a fast-moving field: a further 224 projects are listed by NCBI as underway but yet to be completed. Of these 535 sequenced or partially-sequenced genomes, many are from organisms of concern to human health. Examples of important human pathogens that have been sequenced include *Y. pestis*, the etiological agent of plague (Parkhill *et al.*, 2001); *Neisseria meningitidis*, which causes meningitis (Tettelin *et al.*, 2000); *Vibrio cholerae,* which causes cholera, (Heidelberg *et al.*, 2000); the food-borne zoonosis *Campylobacter jejuni* (Parkhill *et al.*, 2000) and the obligate human pathogen *S. enterica* serovar Typhi (Parkhill *et al.*, 2001). In part, it is

the desire to understand the evolution of pathogens that drives these bacterial genome sequencing projects.

Genetic diversity is caused by either mutation or recombination events, and plasmid transfer. Recombination produces different combinations of alleles as a result of the physical exchange or acquisition (by lateral transfer) of DNA *between* organisms (or between chromosomes for higher organisms). Mutation is the alteration of DNA sequence *within* an organism. A synonymous, or silent, mutation occurs when the altered DNA sequence does not result in a change in the amino acid sequence. A non-synonymous mutation causes a change in amino acid sequence. Non-synonymous mutations include the loss or deletion of DNA, with deletions of just one nucleotide having the potential to cause a change in amino acid sequence. The loss or gain of genes resulting in phenotypic alteration is referred to as 'genome plasticity'. It is important to bear in mind that genome plasticity can be deleterious as well as advantageous to the organism, for example, whilst the loss of genes from the folate biosynthesis pathway makes both *Mycoplasma* spp. and *Treponema* spp. non-susceptible to the antibiotic trimethoprim (Burman, 1986), these organisms are also auxotrophic for folic acid and are therefore forced to utilise exogenous sources of this essential cofactor (de Crecy-Lagard *et al.*, 2007). An example of a deleterious effect of gene deletion in the human genome is Williams-Beuren syndrome, brought about by a deletion in chromosome 7.

It is through a combination of gene loss, acquisition, and mutation that evolution of a genome occurs, driven by the biological and ecological factors that constitute natural selection (Pallen and Wren, 2007). In this way genomes diversify – strains of the same species each share a set of 'core genes' with individual strains possessing extra 'accessory genes' in various combinations (Dorrell, 2005). The rate of evolution of a given species is governed by the rate at which progeny are produced. The relatively fast replication time for bacteria (compared to higher organisms) means that they are able to evolve rapidly in response to their environment.

1.2.3.2. How is comparative genomics data used?

Comparative genomic data can be used to study the evolutionary history of a species, for example a study to investigate the relatedness of *V. cholerae* strains from different cholera biotypes, found < 1% divergence at the genome level for all strains tested (Dziejman *et al.,* 2002). This study also provided important evidence for the way in which *V. cholerae* has evolved into a human pathogen, through the identification of sets of 'core genes' shared by all strains in a species it was possible to identify those genes unique to strains with a particular phenotype, for example colonization of a new (human) host species.

Another use for data generated from comparative genomics studies is in epidemiological surveillance, Kidgell and Winzeler wrote in 2005 that "accurately distinguishing between closely related strains is essential in cases of pathogens associated with food-borne diseases or bioterrorism"

(Kidgell and Winzeler, 2005). This may prove to be particularly important in monitoring the spread of antibiotic resistance and pathogenicity genes, or perhaps in the provision of forensic evidence for legal processes.

Genome comparison between strains of bacteria is also important when contextualising transcriptome analysis as changes in gene expression between strains may result from genomic heterogeneity as well as from differential gene expression.

Molecular techniques used for the analysis of genetic diversity include pulsed-field gel electrophoresis, microsatellite mapping (Dearlove, 2002), mass spectrometry-based genotyping (Pusch *et al.,* 2002), ribotyping (Ansaruzzaman *et al.,* 1996), and restriction fragment length polymorphism mapping. However, whilst these techniques are useful for typing studies, their utility is limited for examination of individual genes, or for ascertaining differences in genetic profile at the strain level. Subtractive hybridisation is the most commonly used method to screen for genes that are additional to an unsequenced strain compared to the sequenced strain (e.g. Soule *et al.,* 2005).

The most thorough method for comparing genomes is of course whole genome sequencing, but although the financial and temporal burdens of this process have become much reduced in recent years, typically only one strain per species (or perhaps subspecies) is sequenced. This limits the utility of

whole genome sequencing for comprehensive comparative genomics studies.

1.2.3.3. Array comparative genomics hybridisations (aCGH)

Although DNA microarrays were originally designed to measure gene expression levels (Lockhart *et al.*, 1996), they also provide a useful platform for evaluating genetic diversity. The higher density the microarray, i.e. the more elements or probes arrayed, the wider the range of potential comparative genomics applications, from strain typing, to polymorphism discovery and even genome sequencing through the use of resequencing arrays (Zwick *et al.,* 2005). An example of a study using high density microarrays is the comparative genomic analysis of *Plasmodium falciparum* which causes malaria. The gDNA from four pathogenic strains was hybridised to an Affymetrix microarray resulting in the discovery of 981 single nucleotide polymorphisms (SNPs), many of which were located in genes associated with varying the antigenic and adhesive character of the parasite (Volkman *et al.,* 2002).

The highest density microarrays are whole genome tiling arrays – non-overlapping or slightly overlapping probes that cover, or tile, the genome from end to end (Mockler *et al.*, 2005). These microarrays are particularly important for comparative genomics studies, as data can be obtained from intergenic regions which are often not included in microarrays that have been designed for expression studies. Changes to intergenic regions do not usually result in change of gene expression or phenotype and are therefore

likely to be tolerated, meaning that intergenic regions tend to display more genome plasticity than coding sequences, making comparison of these regions particularly important in surveillance applications. A more recent use for microarrays designed to include intergenic regions is in a version of chromatin immunoprecipitation analysis known as ChIP-chip or genome-wide location analysis. This is a technique to identify DNA sequences that bind specific DNA-binding proteins such as transcriptional regulators. The DNA sequences that bind proteins can include transcription regulators or other functional elements such as promoters or repressors. A known protein is bound to DNA *in vivo* and the binding protein is cross-linked to the chromatin to which the DNA is attached. The cell is lysed and the DNA inside fragmented by sonication, during which the DNA-binding protein-DNA complex remains intact. This complex is then immunoprecipitated using an antibody targeted to the DNA-binding protein followed by separation of the complex into its protein-DNA components. The DNA region can then identified by hybridisation to a microarray.

Resequencing arrays offer highest resolution and discriminatory power of a DNA microarray to date: four probes are designed for each base pair in a given sequence. One is specific to the reference DNA whilst the other three differ at the central base, coding for each of the three alternative nucleotides. An increase in signal from one of the four probes compared to the three alternatives indicates the correct nucleotide at the central position. This type of array is highly tiled, with every base represented many-fold and this enormously powerful approach can be used to sequence entire genomes

in a single hybridisation experiment. For example, the genomes of several isolates of SARS-Coronavirus were rapidly sequenced in response to the 2004 outbreak, helping to assess the genetic diversity and epidemiology of the disease (Wong *et al.*, 2004).

Phylogenetic trees are plotted from microarray data as follows: DNA from different strains is hybridised to the microarray and the regions of difference (RD) for each strain compared to the reference strain (genome sequenced strain used to design the microarray) are determined. All strains are compared to one another in a pair-wise fashion to see if the RD is present in one or both, or absent from both. The genealogical relationship between strains then becomes apparent (Winzeler *et al.,* 2003).

1.2.3.4. Advantages and limitations of aCGH

Microarrays have several advantages over other methods for comparative genomics studies. The analysis is genome-wide so that large amounts of data can be generated from a single hybridisation reaction. It is possible to design the microarray to extremely high resolution, allowing the detection of fine differences, down to the base pair level. The rapid nature of experiments allows for high throughput analyses, and the stringency and reproducibility of hybridisations reduces the rate of false-positive identification of RDs.

However, there are limitations to the use of DNA microarrays for comparative genomics, the main one being that this platform is asymmetric,

with all data generated relative to the reference strain to which the microarray was designed. Microarrays cannot be used for the detection of additional DNA that has either been acquired by the test strain, or that has been previously lost by the reference strain. In a similar way, diversity resulting from duplications and genome rearrangements is not easy to determine without the use of tiling arrays.

It should be remembered that microarrays with lower discriminatory powers, for example those with one probe per gene, may miss even gross regions of heterogeneity if these fall outside of the region to which the probe is designed. These factors mean that there is a high false-negative rate for detection of RDs using conventional, low density microarrays. However, although tiling, and especially resequencing, arrays have great discriminatory power, they are expensive to build because of the number of probes required. Taking the example of *F. tularensis* Schu S4 which has a genome of 1.89 megabases, predicted to contain 1852 CDS (Larsson *et al.*, 2005): a resequencing array would require 7,200,000 oligonucleotides and a tiling array constructed using 25-mers would require 72,000 oligonucleotides, whereas a conventional microarray designed to investigate gene expression requires just 1,800 probes.

1.2.4. Functional genomics

The aim of a functional genomics, or transcriptomics, study is to examine the function and expression pattern of genes in response to a specified set of conditions. In 2002 G.K. Schoolnik wrote that "the ultimate

goal of whole genome expression studies of pathogenic bacteria is the identification of bacterial genes that are differentially regulated in the host. Within this class of genes are those that adapt the microbe to host-specific microenvironments or encode virulence determinants." (Schoolnik, 2002).

Messenger RNA (mRNA) is isolated under conditions of interest, for example from diseased versus healthy tissue, or from bacteria cultured under nutrient-replete versus nutrient-limited conditions. Reverse transcriptase is used to transcribe RNA into cDNA, with random incorporation of labeled nucleotides. The labeled cDNA is hybridised to the microarray and the test and control expression profiles (transcriptomes) are compared.

1.2.4.1. Advantages and limitations of microarray functional genomics studies

As with aCGH (section 1.2.3.3.), the main advantage of microarrays for functional genomics studies is that the response of a great number of genes to a given condition may be examined in parallel, removing the need for lengthy initial analyses on a gene-by-gene basis.

One of the major potential pitfalls of any study where RNA is quantified as an indication of gene expression is that RNA, and in particular bacterial RNA, is extremely labile (Sarkar, 1997; Taljanidisz *et al.*, 1997). The half-life of 99% of *E. coli* mRNA transcripts have been shown to range from one to 18 minutes (min.), with a mean half-life of five min. depending on the specific growth medium used to culture the bacteria (Bernstein *et al.*,

2002). This enables bacteria to adapt their transcription profiles very rapidly in response to environmental stimuli, and as such, care must be taken not to introduce transcriptional artefacts which could influence the data obtained. Additionally, one of the mechanisms of control of gene expression is by the prevention of translation of mRNAs into a final gene product. This means that the presence of a particular mRNA detected by microarray in the transcriptome does not necessarily indicate that a gene product will be obtained. Also, as mentioned previously, only genes that are present in the organism used to construct the microarray can be detected, meaning that transcription of genes that may have been acquired by strains other than the microarray reference strain will be missed. One example where this limitation could be of particular importance is if antibiotic resistance genes have been acquired by a pathogenic organism as these would not be detectable by a microarray unless it had been constructed with this type of analysis in mind.

1.2.5. Aim

The aim of the work presented in this thesis was to design and construct a *F. tularensis* microarray using the genome sequence of a fully virulent *F. tularensis* subsp. *tularensis* strain, Schu S4. This microarray was tested by performing a comparative genomics study and comparing the data obtained to those reported from previously published comparative genomics studies carried out using *Francisella*. The comparative genomics data obtained in this study were also analysed for clues as to the genetic basis of the attenuation of *F. tularensis* LVS, and of the relative attenuation of *F.*

tularensis subsp. *holarctica* compared to *F. tularensis* subsp. *tularensis*. Finally, the microarray was used to perform functional genomics studies using *F. novicida* cultured under conditions selected to represent some of the conditions thought to be encountered by *Francisella in vivo*. Data from both aCGH and functional genomics studies were ultimately utilised in a wider *F. tularensis* vaccine programme, in the provision of target genes for an improved rationally attenuated vaccine strain.

Chapter 2

Methods

2.1. Bacterial strains and cultivation

Bacterial strains used in this study are listed in table 2.1. *F. tularensis* was cultured on blood glucose cystine agar (BCGA) (table 2.2), or in Chamberlain's chemically defined medium (CDM) (table 2.3). Agar plates were inoculated from other plates, or from stocks held at -80°C in 40% glycerol made up with phosphate buffered saline (PBS) (Gibco, Paisley UK). Plates were incubated, inverted, at 37°C for 24-48 hours (h), until growth was observed. Broth cultures were incubated at 37°C (with the exception of cultures under elevated temperature) with shaking at ~200 revolutions per minute (rpm). All *F. tularensis* subsp. *tularensis* and *F. tularensis* subsp. *holarctica* manipulations were carried out at ACDP level 3, with agar plate incubation carried out in secondary containment. All *F. novicida* manipulations were carried out at ACDP level 2, with agar plate incubation carried out in secondary containment and broth incubations carried out in disposable 250 ml Erlenmeyer flasks with filtered lids (Nalge Nunc International, New York, USA).

2.2. Isolation and quantification of DNA from *Francisella*

DNA from *Francisella* strains was isolated using the PUREGENE DNA isolation kit (Gentra Systems, Minneapolis, USA) following the manufacturer's instructions. Bacteria were cultured on BCGA as described in section 2.1., and ~ 0.5 to 1.5 billion cells (a loopful) was scraped from the agar and suspended in 500 microlitres (µl) PBS. The suspension was centrifuged at 16,000 x *g* for one minute (min.). The supernatant was discarded and the bacterial pellet was resuspended in 300 µl PUREGENE

Cell Lysis Solution. The **Table 2.1** Bacterial strains used in this study. Strain origin and alternative designation details taken from Svensson *et al.*, (2005). [a]Isolated from *F. tularensis* FSC074. [b]Isolated from *F. tularensis* FSC158.

Strain Number	Subsp.	Date	Location	Source	Alternative designation
FSC 237	*tularensis*	1941	Ohio, USA	Human ulcer	Schu S4
FSC 043	*tularensis*	1941	Ohio, USA	Human ulcer	Schu
FSC147	*mediaasiatica*	1965	Kazakhstan	Miday gerbil	GIEM 543
FSC 124	*holarctica*	1990	Ukraine, Odessa region	Water	14588
FSC 257	*holarctica*	1949	Moscow area	Tick	503/840
FSC 012	*holarctica*	1941	Montana, USA	Tick	425F4G
FSC 200	*holarctica*	1998	Ljusdal, Sweden	Human	
FSC352	*holarctica*	1974	Nas, Sweden	Hare[a]	
FSC354	*holarctica*	1974	Nas, Sweden	Hare[a]	
FSC358	*holarctica*	1994	Norway	Human[b]	
FSC 338	*holarctica*		Russia	Vaccine strain	Strain 15
FSC 155	*holarctica*		Russia	Vaccine strain	LVS, ATCC 29684
FSC 040	*novicida*	1950	Utah, USA	Water	ATCC 15482

Table 2.2 BCGA (Atlas, 1993)

Reagent	Concentration (% v/v)
Nutrient agar (1.5% (w/v) agar, 0.05% (w/v) pancreatic digest of gelatine, 0.3% (w/v) beef extract, Autoclaved 15 min at 15 psi, 121°C)	85
Glucose cystine solution (25% (w/v) glucose, 1% (w/v) cystine·HCl, Filter sterilized)	10
Defibrillated horse blood	5

Table 2.3 CDM (Chamberlain, 1965)

Reagent	Concentration (% w/v)
L-arginine(free base)	0.04
L-aspartic acid	0.04
L-cysteine·HCl	0.02
L-histidine (free base)	0.02
DL-isoleucine	0.04
L-leucine (methionine free)	0.04
L-lysine (mono HCl)	0.04
DL-methionine	0.04
L-proline (hydroxyl-L-proline-free)	0.20
DL-serine	0.04
DL-threonine (allo-free)	0.20
L-tyrosine	0.04
DL-valine	0.04
Spermine phosphate	0.04
Thiamine·HCl	0.0004
DL-calcium pantothenate	0.002
Glucose	0.40
NaCl	1.0
$MgSO_4 \cdot 7H_2O$	0.0135
$FeSO_4 \cdot 7H_2O$	0.0002
KH_2PO_4	0.10
K_2HPO_4	0.10
pH6.2 – 6.4, filter sterilized	

sample was incubated at 80°C for 5 min. to lyse the cells. After incubation, 1.5 µl PUREGENE RNaseA was added to the cell lysate and mixed by inversion 25 times. The sample was incubated at 37°C for 1 h. The sample was cooled to room temperature and 100 µl PUREGENE Protein Precipitation Solution was added. The sample was vortexed for 20 seconds (s.), and then centrifuged at 16,000 x *g* for three min. to pellet the precipitated proteins. The supernatant containing the DNA was transferred to a fresh 1.5 millilitre (ml) tube containing 300 µl 100% isopropanol. The sample was gently inverted 50 times to precipitate the DNA and was then centrifuged at 16,000 x *g* for one min. to sediment the DNA. The supernatant was discarded and 300 µl 70% ethanol was added. The sample was inverted several times to wash the DNA and then centrifuged at 16,000 x *g* for one min. The supernatant was discarded and the DNA was air-dried for ~10 min., until all visible liquid had evaporated. The pellet was resuspended in 50 µl PUREGENE DNA Hydration Solution and incubated at 65°C for 1 h with occasional inversion to aid dispersal of the pellet. DNA quantification was performed by spectrophotometric analysis of 1 µl aliquots using an ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, USA).

2.2.1. Concentration of DNA by ethanol precipitation

The following were added to each DNA sample: 1/10 volume ice cold sodium acetate (3 molar [M], pH 5.2), 2 volumes ice cold absolute ethanol. The samples were placed at –20°C for a minimum of 1 h and then centrifuged at 16,000 x *g* for 30 min to sediment the precipitated DNA. The supernatants were decanted and 0.75 ml 70% ethanol was added to each

sample. The samples were centrifuged as above for 5 min. The supernatants were decanted and the DNA pellets allowed to air-dry for a maximum of 10 min. Each DNA pellet was resuspended in nuclease-free water (Ambion, Austin, USA) to a final concentration of 1 mg ml^{-1}. Samples were placed at 65°C for 30 min to aid resuspension of the DNA. DNA samples were stored at –20°C until use.

2.3. PCR

2.3.1. PCR primer storage

PCR primers were supplied lyophilised (MWG-Biotech, Ebersberg, Germany) and were resuspended in nuclease-free water to a concentration of 100 picomole per microlitre (pmol μl^{-1}), according to the individual molecular weight supplied with the primers. A working dilution of 10 pmol μl^{-1} of each primer was made and stored at 4°C. The primers at 100 pmol μl^{-1} were stored at -20°C.

2.3.2. PCR reaction

The following were added to 0.2 ml thin-walled nuclease-free reaction tubes (Ambion):

illustra PuReTaq Ready-To-Go PCR bead (GE LifeSciences, Amersham, UK)	1
DNA at 1 micrograms (µg) ml^{-1}	1 µl
Forward primer at 10 pmol μl^{-1}	2 µl
Reverse primer at 10 pmol μl^{-1}	2 µl
Nuclease-free water (Ambion)	20 µl

Reactions were performed in a GeneAmp PCR System 2700 (Applied Biosystems (ABI) (Warrington, UK) thermalcycler as follows:

Stage	Temperature (°C)	Time	Replicates
1	95	5 min.	1
2	95 Primer T_m - 5°C 72	30 s. 30 s. 1 min.	30
3	72	10 min.	1
4	10	∞	1

Reactions were resolved using an agarose gel and visualised using ethidium bromide, with the concentration of agarose and choice of molecular weight marker being dependent on the expected amplicon size: amplicons of ≤ 500 base pairs (bp) were resolved using agarose at 2% (w/v) dissolved in Tris-acetate-ethylenediaminetetraacetic acid (TAE) buffer (Sigma, Bournemouth, UK) with 50 bp ladder DNA marker (GE LifeSciences), and amplicons of ≥ 500 bp were resolved using 1% (w/v) agarose with KiloBase DNA marker (GE LifeSciences). Agarose gels were visualised using an ultra violet transilluminator.

2.4. Culture of *F. novicida* under *in vitro* stress conditions

2.4.1. Culture of *F. novicida* under iron-depleted conditions

F. novicida was cultured at 37°C in 10 ml CDM for ~ 16 h with shaking at ~190 rpm to give a stationary phase culture. The overnight suspension was diluted in pre-warmed CDM (2 ml in 28 ml) and incubated at 37°C, with shaking at ~190 rpm for 2.5 h so that the culture was in logarithmic (log) phase. The culture was centrifuged at 10,000 x *g* for 10 min at 4°C to sediment the cells. The supernatant was decanted and the cells resuspended in 30 ml PBS. This PBS wash was repeated three further times to remove any residual CDM. The washed bacterial suspension was added (2 ml) to 28 ml pre-warmed CDM (control) and to CDM without $FeSO_4 \cdot 7H_2O$ but supplemented with 10 millilmolar (mM) deferroxamine (CDM-Fe). The control culture contained iron at a concentration of 2 x 10^{-4} % (w/v). The growth rate of test and control cultures was compared by measuring the absorbance at 260 nanometres (A_{260}) of each at hourly intervals. Both test and control cultures were grown at 37°C in an air-incubator, pH 6.2 with shaking at 190 rpm: this was T_0.

2.4.2. Culture of *F. novicida* in CDM without methionine

F. novicida was cultured at 37°C in 10 ml CDM for ~16 h with shaking at ~190 rpm to give a stationary phase culture. The overnight suspension was diluted in pre-warmed CDM (2 ml in 28 ml) and incubated at 37°C, with shaking at ~190 rpm for 2.5 h so that the culture was in log phase. The culture was centrifuged at 10,000 x *g* for 10 min at 4°C to sediment the cells. The supernatant was decanted and the cells resuspended in 30 ml PBS.

This PBS wash was repeated three further times to remove any residual CDM. The washed bacterial suspension resuspended in 30 ml pre-warmed CDM made with DL-methionine omitted (CDM-met).

2.4.2.1 Culture of *F. novicida* in CDM-met under low pH

F. novicida was cultured as described in section 2.4.2, in duplicate. The pH of one of the flasks of CDM-met was adjusted from pH 6.2 to pH 5.4 by the addition of 100 µl hydrochloric acid (HCl) (1 M), before the bacteria were added. The second (control) culture was maintained at pH 6.2. Both cultures were incubated at 37°C, with shaking at ~190 rpm – this was T_0. The growth rate of test and control cultures was compared by measuring the absorbance at 260 nanometres (A_{260}) of each at hourly intervals.

2.4.2.2 Culture of *F. novicida* in CDM-met under elevated temperature

F. novicida was cultured as described in section 2.4.2, in duplicate. One of the cultures was incubated at 42°C with shaking at ~190 rpm. The second (control) culture was incubated at 37°C with shaking at ~190 rpm – this was T_0. The growth rate of test and control cultures was compared by measuring the A_{260} of each at hourly intervals.

2.4.2.3. Culture of *F. novicida* in CDM-met under oxidative stress

F. novicida was cultured as described in section 2.4.2, in duplicate. Hydrogen peroxide (H_2O_2) (150 µl at 1 M) was added to one of the flasks of CDM-met to give a final concentration of 5 mM, before the bacteria were added. No H_2O_2 was added to the second (control) culture. Both cultures

were incubated at 37°C, with shaking at ~190 rpm – this was T_0. The growth rate of test and control cultures was compared by measuring the A_{260} of each at hourly intervals.

2.5. RNA isolation

2.5.1. Stabilisation of RNA from *F. novicida* cultured *in vitro*

At intervals ~4 x 10^9 colony forming units (CFU) *F. novicida* was removed from each culture and immediately added to two volumes RNAProtect Bacteria Reagent (Qiagen, Crawley, UK). The cell/RNAProtect mixture was shaken and left at room temperature for 5 min. The samples were then stored at 4°C for up to 2.5 h before RNA isolation.

2.5.2 RNA extraction and quantification from *F. novicida* cultured *in vitro*

RNA from *F. novicida* was extracted using the RNeasy Midi Kit (Qiagen) following the manufacturer's instructions. Briefly, this involved sedimenting the bacteria by centrifuging at 10,000 x *g* for 10 mins. The bacteria were lysed using egg white lysozyme (1 milligram [mg] ml^{-1}). Buffer RLT (Qiagen) containing β-mercaptoethanol (10 microlitres [μl] ml^{-1}) was added, followed by absolute ethanol. The sample was applied to an RNeasy midi column (Qiagen) in a 15 ml collection tube and centrifuged at 5000 x *g* for 5 min. The eluate was discarded and buffer RW1 (Qiagen) was applied to the column. The column was centrifuged at 5000 x *g* for 5 min and the eluate was discarded. Buffer RPE (Qiagen) was applied to the column which was then centrifuged at 5000 x *g* for 2 min. The eluate was discarded

and buffer RPE was applied to the column again. The column was centrifuged at 5000 x *g* for 5 min and the eluate was discarded. The column was transferred to a fresh, RNase-free, 15 ml collection tube and RNase-free water was added to the column. The column was allowed to stand for 1 min and then centrifuged at 5000 x *g* for 3 min to elute the sample. All samples were stored at –20°C prior to quantification and concentration. RNA quantification for each sample was performed by spectrophotometric analysis of 1μl aliquots using an ND-1000 spectrophotometer.

2.5.3. Concentration of RNA by ethanol precipitation

Concentration of RNA samples was carried out as described in section 2.2.1., but concentrated RNA was stored at -80°C until use (whereas concentrated DNA was stored at -20°C).

2.6. Microarrays

2.6.1. Printing *F. tularensis* DNA microarrays

DNA microarray design and printing is described in chapter 3.

2.6.2. CyDye labeling of DNA

2.6.2.1. Method 1, gDNA labeling

This method is based on that published by Stabler *et al.*, (2003), and Stewart *et al.*, (2002). One reaction was prepared per sample, per microarray slide. The following reagents were combined in an amber 1.5 ml microcentrifuge tube (the tube was amber to block out light and help to prevent photo-bleaching of the CyDyes):

F. tularensis Schu S4 gDNA (1 mg ml^{-1})	6.0 µl
Random primers (3 µg/µl (Invitrogen Corporation, Paisley, UK)	1.0 µl
DEPC treated water (Ambion, Huntingdon, UK)	34.5 µl

The reaction was heated at 95°C for 5 min; snap-cooled on ice, and briefly centrifuged to bring the reactants to the bottom of the tube.

The following reagents were added to each reaction:

10x React2 buffer (Invitrogen)	5.0µl
dNTPs (Invitrogen) (5mM dA/G/TTP, 2 mM dCTP)	1.0 µl
Cy3- or Cy5-dCTP Fluorolink	1.5 µl
(GE Life Sciences)	
DNA polymerase I large (Klenow) fragment (Invitrogen)	1.0µl

The reactions were incubated in the dark for 90 min at 37°C.

2.6.2.2. Method 2, cDNA labeling

One reaction was prepared per sample, per microarray slide. The following reagents were combined in an amber 1.5 ml microcentrifuge tube:

RNA (1 mg/ml)	6.0 µl
Random primers (3 µg/µl, Invitrogen)	1.0 µl
DEPC treated water (Ambion)	4.0 µl

The reaction was heated at 95°C for 5 min; snap-cooled on ice, and briefly centrifuged to sediment the reactants.

The following reagents were added to each reaction:

5 x First Strand Synthesis Buffer (Invitrogen)	5.0µl
0.1 M DTT (Invitrogen)	2.5 µl
dNTPs (5mM dA/G/TTP, 2 mM dCTP)	2.3 µl
Cy3- or Cy5-dCTP Fluorolink	1.7 µl
CyScribe reverse transcriptase (GE Healthcare)	2.5 µl

The reactions were incubated in the dark for 10 min at 25°C, and then for 90 min at 42°C.

2.6.3. Purification of CyDye-labeled DNA using MinElute (Qiagen)

Where applicable (i.e. when a two channel hybridisation was to be performed), Cy3- and Cy5-labeled DNA/cDNA were combined in a single microcentrifuge tube. DNA purification was then carried out using a MinElute Kit, following the manufacturer's instructions. Briefly, five volumes of buffer PB (Qiagen) was added to the sample and mixed by pipetting. The sample was applied to a MinElute column (Qiagen) in a 2 ml collection tube. The column was centrifuged at 16,000 x *g* for 1 min and the eluate was discarded. Buffer PE (750 µl) was applied to the column and centrifuged as before. The eluate was discarded; the column was replaced into the same 2 ml collection tube and centrifuged as before. The column was placed into a fresh 1.5 ml collection tube and 10 µl sterile water was applied directly to the

column membrane. The column was allowed to stand for 1 min and then centrifuged as before to elute the labeled DNA.

2.6.4. Hybridisation

Purified CyDye-labeled DNA (10 µl) was mixed with 10 µl 1x DIG Easy Hyb buffer (Roche, Lewes, UK) and heated at 70°C for 2 min. The sample was cooled to room temperature for 3 min and then briefly centrifuged to sediment the reactants.

The microarray was placed into a hybridisation chamber (Genetix, New Milton, UK) and 4 ml water was added to the well at the bottom. A 22 x 22 millimetre (mm) glass LifterSlip (Eerie Scientific, Portsmouth, USA) was gently placed onto the microarray, to cover the printed area. Care was taken to ensure that the LifterSlip bars were face down, providing a space for the target DNA sample to be applied.

The labeled target DNA solution was carefully pipetted under the LifterSlip, allowing the solution to be drawn under and across the array by capillary action. The hybridisation chamber was sealed tightly and incubated in a dry oven at the desired hybridisation temperature.

2.6.5. Washing

Wash buffers FT1 (table 2.4) and FT2 (table 2.5) were pre-warmed to 42°C overnight. Hybridised microarrays were washed for 30 min in low stringency buffer FT1, with the buffer refreshed after 15 min. Microarrays

were then washed for 30 min in high stringency buffer FT2, again with the buffer refreshed after 15 min. Washes were carried out using a microarray 10 slide wash rack (Genetix) in a 1 litre (l) beaker. Washes were carried out in the dark, at room temperature (but using buffers pre-warmed to 42°C) using a magnetic stirrer to gently agitate the buffers. After the second FT2 wash, the microarrays were briefly rinsed in isopropanol and then dried by centrifugation at 215 x *g* for 5 min.

Table 2.4 Buffer FT1

Reagent	Concentration
20x sodium chloride-sodium citrate buffer (SSC) (3 M NaCl, 0.3 M Na citrate [pH7.0]) (Sigma)	10% (v/v)
Sodium dodecyl sulphate (SDS) (10% w/v solution)	5% (v/v)
Distilled H_20	85% (v/v)

Table 2.5 Buffer FT2

Reagent	Concentration
20x SSC	1.25% (v/v)
SDS (10% w/v solution)	5% (v/v)
Distilled H_20	93.75% (v/v)

2.6.6. Scanning and image quantification

Hybridised and washed microarrays were imaged using the GenePix 4000B scanner (Molecular Devices Corporation, Sunnyvale, USA), controlled by GenePix Pro software (Molecular Devices Corporation). Images were quantified using BlueFuse v.3.1 software (BlueGnome, Cambridge, UK) using the auto-gain function.

2.6.7. Data acquisition and analysis for transcriptomic studies

Fluorescent scanning and image quantification was carried out as described in section 2.6.6., using a GenePix 4000B scanner and BlueFuse v.3.1 image analysis software. Quality control was performed by generating a normalisation constant for each array by excluding positive and negative control gene spots from each data set, as well as those with confidence values <0.1 and taking the median of the remaining $\log_2$ ratios between channels. Linear normalisation was then applied by subtracting the normalisation constant from each $\log_2$ ratio. After global scaling, the relative expression for each time point was determined using the $\log_2$ ratio from each paired test and control data set. To test statistically whether each gene was ≥ 1.5-fold regulated a significance p-value was determined using a two-tailed t-test. The Benjamini Hochberg procedure for controlling the false discovery rate (FDR) was applied to generate a corrected p-value (P_{FDR}).

2.7. Quantitative real time PCR (QPCR)

TaqMan assays were designed and synthesised by ABI, using their proprietary Assays-by-Design service. Each assay included a FAM dye-

labeled TaqMan minor-groove binder probe and two unlabeled PCR primers specific to the gene of interest. Unlabeled cDNA was prepared as described in sections 2.6.2.2. and 2.6.3., using an equal mix of all four dNTPs (5 mM each) and without using Cy-dCTP. The following were added to per well of a 96-well optical reaction plate (ABI):

TaqMan Universal Mastermix (ABI)	12.5 µl
TaqMan Assays-by-Design bespoke assay (ABI)	1.25 µl
DEPC-treated water	9.25 µl
Unlabeled cDNA	2.0 µl

The plate was covered with an Optical Adhesive Cover (ABI) and was briefly centrifuged to sediment the contents. A heat proof cover designed to prevent evaporation from 96-well plates (ABI) was applied to the plate. The PCR was carried out using a Prism 7000 (ABI), using Prism 7000 sequence detection system software (ABI) under the following cycling conditions:

Stage	Temperature (°C)	Time	Replicates
1	50	2 min	1
2	95	10 min	1
3	95 60	15 s 1 min	40

Relative quantification of gene expression between test and control samples was calculated as follows: a threshold level was set during the exponential phase of amplification, and the cycle threshold (CT) of each reaction was measured from the cycle number at which the reaction signal corresponded with the threshold level. Each test and control assay was carried out in triplicate, and for each assay non-template controls were also included in triplicate. For each test condition, a 'housekeeping' gene, reported to have a stable expression pattern was also assayed. Quantitation of gene expression was calculated using the 2(-delta delta C(T)) ($\Delta\Delta$CT) method (Livak and Schmittgen, 2001). Briefly, the mean of the CT values for both test and control assays was taken, both for the gene of interest and for the housekeeping gene. The delta CT (ΔCT) was calculated by subtracting the mean CT of the housekeeping gene from the mean CT of the test gene. The $\Delta\Delta$CT was then calculated by subtracting the ΔCT of the housekeeping gene from the ΔCT of the test gene. A housekeeping gene is so named because the level of expression is assumed to remain the same regardless of environmental stimuli. Any differences between the measured level of expression of the housekeeping gene in test and control samples are therefore attributed to variance in the amounts of RNA template in test and control cultures, and results for test genes adjusted accordingly. This is a controversial concept, especially when considering relatively small bacterial genomes as it seems doubtful that any given gene would not be affected by some external stimuli. With this in mind, the level of change in expression of the housekeeping gene selected (based on previous publication) was checked in the microarray data to be confirmed. If the housekeeping gene

appeared to be regulated in response to the condition under test, then another housekeeping gene would be selected.

Chapter 3

Design and construction of a *F. tularensis* microarray

3.1. Introduction

As described in section 1.2., DNA microarrays are a very powerful tool that can be used to investigate the genetic makeup or transcriptome profile of an organism. DNA probes are arrayed onto a solid platform and labeled DNA targets are hybridised to the probes under stringent conditions. Unbound targets are removed by washing, and the resulting hybridised microarray is viewed under conditions which allow visualisation of labels attached to the bound DNA targets. Two main microarray formats are available commercially, based on glass slides and silicon chips. Silicon chip technology is usually considered the gold-standard, but is very expensive, which means that glass-slide format microarrays are also popular. There are a number of variables in format to be considered at the design stage of glass-slide microarrays, not least whether probes should consist of PCR amplicons or synthetic oligonucleotides.

3.1.1. A *F. tularensis* ordered clone microarray

A *F. tularensis* microarray has previously been generated using PCR products derived from a pUC18 clone library synthesised for the genome sequencing of strain Schu S4 (Broekhuijsen *et al.* 2003). Samples of DNA isolated from 27 *Francisella* strains were hybridised to this microarray in experiments to reveal the extent of gross genetic diversity between strains. The results suggested that there is only limited genetic variance between strains, despite differences in virulence between subspecies (Broekhuijsen *et al.*, 2003). However, there were some limitations to the microarray used: the clones used as PCR substrates to generate probes were generated by

inserting random *F. tularensis* Schu S4 DNA fragments into the pUC18 vector. This means that some of the probes contained more than one CDS or truncated CDS, making it difficult to assign a putative gene name or function to each probe. This limits the value of the ordered clone microarray, particularly for transcriptomic analysis. In addition, the plasmid probes were in limited supply, a second factor highlighting the requirement for an improved version of the *F. tularensis* microarray.

3.1.2. Aim

The aim of this work was to design and construct a second-generation *F. tularensis* microarray, with probe design based on the *F. tularensis* Schu S4 genome sequence. A glass slide format was chosen, with one oligonucleotide probe designed to represent each putative CDS identified in the *F. tularensis* Schu S4 genome sequence.

3.2. Methods

This section describes the methods that were common to all of the hybridisations described in the optimisation of the second-generation microarray, including design of the oligonucleotide probes. Each experiment performed to optimise the microarray hybridisation and wash conditions differed from the last in some aspect as the procedure was refined. However, the DNA labeling reactions and hybridisation methods, apart from hybridisation buffer and temperature which were two of the parameters subject to optimisation, remained the same.

3.2.1. Oligonucleotide design

Oligonucleotides were designed using the *F. tularensis* Schu S4 genome sequence, available at the time as six contiguous pieces (contigs.). The contigs. were concatenated to give a complete sequence and putative Open reading frames (ORFs) were assigned using Glimmer software (Delcher *et al.*, 1999). Concatenation was carried out with the addition of stop codons in all six reading frames between each contig., and before the first contig., in order to prevent the formation of false ORFs. Glimmer was used to find putative ORFs with a minimum gene length of 500 base pairs (bp), of which there were 1282 identified. Putative ORFs were allowed to overlap by 30 bp or 10% maximum. A Hidden Markov Model of the long ORF sequences was built, which was utilised by Glimmer2 (Delcher *et al.*, 1999) to find additional putative CDS of a minimum gene length of 90 bp, which increased the total number of putative CDS identified to 2057. The putative CDS list was analysed against itself using BLASTn (Altschul *et al.*, 1990), and against the NCBI database using tBLASTn (Altschul *et al.*, 1990).

Duplicated CDS and CDS that were thought to be incorrect were deleted, as were CDS shorter than 120 bp in length. The final number of putative CDS used for oligonucleotide design was 1937.

A FASTA file of putative CDS was sent to MWG Biotech, along with the concatenated genome sequence for oligonucleotide design using their proprietary Oligos4Array software. The algorithm was designed to generate several 50-mer complementary oligonucleotides to each CDS. The optimum oligonucleotide for each CDS was selected on the basis of lowest *in silico* level of cross-reactivity with other CDS (using BLASTn and Smith-Waterman alignment searches), and the highest *in silico* binding activity between the oligonucleotide probe and its target DNA. Additional selection criteria included minimum tertiary structure formation, GC content and melting temperature (T_m) falling within given parameters wherever possible, to give a more uniform binding capacity between oligonucleotides. In addition, MWG Biotech were requested to cross-check each oligonucleotide for levels of cross reactivity with the human, mouse and amoeba, as these were the infection models in use, and it was thought likely that future target RNA samples could be contaminated with RNA from these sources. Oligonucleotides were designed using only the central 60% of each CDS, in order to account for possible discrepancies in the precise location of start and stop codons.

3.2.1.1. Control probes

Twenty control CDS were also selected from ten organisms that are unrelated to *F. tularensis* (table 3.1). Local alignment searches were carried out between these CDS and the *F. tularensis* genome to ensure that there was no identity between them. Oligonucleotides were designed to each of these CDS as described above (section 3.2.1.), with the exception that they were not included in the database searches against human, mouse and amoeba genomes.

3.2.1.2. Oligonucleotide format

Oligonucleotides were supplied, un-modified and 'ready-to-spot', at a concentration of 100 pmol μl^{-1} (+/- 20%) in 3 x SSC buffer.

3.2.2. Microarray printing

3.2.2.1. FOI-printed microarrays

In the first instance microarrays were printed at FOI, Umea, Sweden using an SDDC/C200 Robot (Engineering Services Inc., Toronto, Canada). Oligonucleotides were used, as supplied by MWG Biotech Ltd., at 100 pmol μl^{-1} in 3 x SSC buffer. In addition to negative control probes, negative control spots, either empty or 20 x SSC only, were included. Microarrays were printed on CMT-GAPS slides (Corning Inc. Life Sciences, Lowell USA), which were supplied coated with aminosilane. Microarrays were re-hydrated by placing them, face down, over a vessel filled with 1 x SSC buffer until the spots could be seen to glisten. Care was taken not to allow spots to become so wet that they merged. The microarrays were then snap-dried by placing

Table 3.1 Negative control oligonucleotides included on the second-generation *F. tularensis* Schu S4 microarray.

Organism	Gene	Accession No.
A. castellanii	Actophorin	M93361
Naegleria fowleri	mp2Cl5	AY049749
Homo sapiens	Thioredoxin reductase	X91247
A. castellanii	TPBF	L46867
Geobacillus stearothermophilus	*bstYIM*	AY97779
A. castellanii	*rpb1*	U90211
Mus musculus	β-actin	X03672
Dictyostelium discoideum	*tor*	AY204354
M. musculus	*Il28*	NM177396
Shigella flexneri	*mxiM*	M98391
M. musculus	α-tubulin	M13446
Shigella sonnei	*mxiJ*	D50601
M. musculus	*rdh1*	AY072301
Y. pestis	*lcrG*	M26405
H. sapiens	BPIL1	AF465765
Hordeum vulgare	*fp3*	AJ549119
H. sapiens	L37a	L06499
S. enterica serovar Typhimurium	*sinI*	AF140550
H. sapiens	APCDDI	AB104887
Burkholderia cepacia	*bcscC*	AY166598

them, face up, onto an 80°C heat block for 3 seconds. A UV Stratalinker (Strategene, La Jolla USA) set at 500 milli-Joules was used to cross-link the DNA to the slide. Microarrays were blocked in succinic anhydride (0.02 M) and sodium borate (0.02 M) in 1-methyl-2-pyrrolidone (0.08 M), pH 8.0. Blocking was carried out at room temperature for 20 min., with gentle agitation using an orbital shaker. Microarrays were washed in distilled water at 95°C for two min., and then submerged in 95% ethanol for final rinsing and drying by centrifugation at 215 x *g* for seven min. at room temperature.

3.2.2.2. HPA-printed microarrays

Subsequently, a Biorobotics TAS Microgrid II Arrayer (Genomics Solutions Ltd., Huntingdon, U.K.) was used to print microarrays at HPA, Porton Down, Salisbury, UK. The microarray was designed to contain 16 sub-arrays (in a four by four layout), with each sub-array containing 306 probes (in a 17 by 18 arrangement), giving a total of 4896. Oligonucleotides, supplied by MWG Biotech Ltd. at 100 pmol μl^{-1} (+/- 20%), were diluted 1:1 using 50% glycerol. Oligonucleotides corresponding to each *Francisella* CDS were arrayed in duplicate, and the whole microarray was printed onto each slide in duplicate. Negative control oligonucleotides were arrayed at the same frequency as test oligonucleotides i.e. four times per slide. Additional negative controls were included in the form of buffer-only spots, printed using 1 x SSC, and empty, un-printed, spots. Cy3-dCTP was included in the corner of each sub-array to aid with visual alignment of Imagene quantification grids. Microarrays were printed on aminosilane-coated Nexterion A slides (Schott UK Ltd., Stafford, UK). Microarrays were

not chemically post-processed, but the oligonucleotides were fixed to the slides by baking in a dry oven at 180°C for two hours (as opposed to UV crosslinking). Processed microarrays from were stored in a dark, dust free environment with desiccation until use. It was judged that microarrays would remain usable for up to six months after printing.

3.2.3. Preparation of Cy3-labeled DNA and hybridisation

DNA was isolated from *F. tularensis* Schu S4 as described in section 2.2. The DNA was labeled with Cy3-dCTP (6 µg DNA per reaction) as described in section 2.6.2.1. and purified as described in section 2.6.3. The labeled DNA was hybridised to the microarray as described in sections 2.6.4. to 2.6.6. The hybridisation temperature and buffer conditions were varied as part of the microarray optimisation as detailed in section 3.3.

3.2.4. Microarray washing

3.2.4.1. Wash protocol 1

Wash protocol 1 was carried out at room temperature with agitation.

Buffer	Composition in water	Length of wash (min.)
Wash 1	1 x SSC, 0.2% SDS	1
Wash 1	1 x SSC, 0.2% SDS	2 (x 2)
Wash 2	0.1 x SSC, 0.2% SDS	1
Wash 2	0.1 x SSC, 0.2% SDS	2
Wash 3	0.1 x SSC	1 (x 2)
Wash 4	isopropanol	1 (submerged without agitation)

3.2.4.2. Wash protocol 2

Wash protocol 2 was carried out at 42°C, using buffers pre-warmed to 42°C. Washes were carried out in the dark in a rotisserie oven, set to the lowest rpm of 1.

Buffer	Composition in water	Length of wash (min.)
FT1	2 x SSC, 0.5% (w/v) SDS	15 (x 2)
FT2	0.5 x SSC, 0.5% (w/v) SDS	15 (x 2)

3.2.4.3. Wash protocol 3

Wash protocol 3 was carried out at 42°C, using buffers pre-warmed to 42°C. Washes were carried out in the dark in a rotisserie oven, set to the lowest rpm of 1.

Buffer	Composition in water	Length of wash (min.)
FT1	2 x SSC, 0.5% (w/v) SDS	15 (x 2)
FT3	0.25 x SSC, 0.5% (w/v) SDS	15 (x 2)

After washing microarrays were dried by centrifugation in Sorvall Legend M bench top centrifuge at 215 x *g* for 7 minutes.

3.2.5. Scanning and image quantification

Microarrays were scanned at 550 nm using a 428 scanner (Affymetrix, Inc., Santa Clara, California, USA) and quantified using Imagene v.5.5 (Biodiscovery, Inc., El Segundo, USA). The median background was subtracted from the median signal for each spot to give a true (background corrected) value.

3.3. Results

All microarray hybridisation and washing optimisation steps were carried out using *F. tularensis* Schu S4 gDNA. As the oligonucleotides were complementary to the *F. tularensis* Schu S4 genome sequence, each test probe was expected to give a high signal, and negative control probes were expected to give low or no signals after hybridisation *with F. tularensis* Schu S4 DNA.

3.3.1. Microarrays printed at FOI

A total of 25 microarrays were printed at FOI. A summary of the optimisation steps taken with the FOI-printed microarrays detailed below can be found in table 3.2. Ten experiments were performed to optimise the hybridisation and washing conditions used with this microarray, however it eventually became clear that a systematic fault with the microarrayer was causing cross-contamination of the oligonucleotide probes, evidenced by positive hybridisation signals from negative control spots (data not shown).

3.3.2. Microarrays printed at HPA

A total of 50 microarrays were printed at HPA in the first instance. A summary of the optimisation steps taken with the HPA-printed microarrays detailed below can be found in table 3.3.

Table 3.2 Overview of hybridisation optimisations carried out using FOI-printed microarrays.

Hyb. conditions	Hyb. temp.	Wash conditions	Comments	Result
Low salt (DIG Easy Hyb)	42°C	Protocol 1	Performed at FOI using Lucidea hyb. station.	High signals from controls
High Salt	61°C	Protocol 1		Test probes variable. High signals from controls
Low salt (DIG Easy Hyb)	42°C	Protocol 1	Performed at Dstl without using Lucidea hyb. station.	Test oligos variable. High signals from controls
Low salt (DIG Easy Hyb)	65°C	Protocol 1		No signal
Low salt (DIG Easy Hyb)	70°C	Protocol 1		No signal
Low salt (DIG Easy Hyb)	42°C	Protocol 2		Test probes variable. High signals from 5% of controls
Low salt (DIG Easy Hyb)	42°C	Protocol 2	30 minute wash steps	Test probes variable. High signals from 20% of controls
Low salt (DIG Easy Hyb)	42°C	Protocol 2		No signal – unclear why
Low salt (DIG Easy Hyb)	42°C	Protocol 3	30 minute wash steps	Test probes variable. High signals from 10% of controls
Low salt (DIG Easy Hyb)	42°C	Protocol 3		Apparent that false positives attributable to cross contamination

Table 3.3 Overview of hybridisation optimisations carried out using HPA-printed microarrays.

Hyb. conditions	Hyb. temp.	Wash conditions	Comments	Result
Low salt (DIG Easy Hyb)	42°C	Protocol 3	First HPA print run	Fewer false positives than seen before, no evidence of print tip carry over
Low salt (DIG Easy Hyb)	42°C	Protocol 3	Microarray Blocked with salmon DNA	Hyb. failed
Low salt (DIG Easy Hyb)	42°C	Protocol 3	Second HPA print run using fresh oligos	~0.4 % false positive rate. Even test signals
Low salt (DIG Easy Hyb)	42°C	Protocol 3		~2.2 % false positive rate. Even test signals
Low salt (DIG Easy Hyb)	42°C	Protocol 3		1.5 % false positive rate. Even test signals

3.3.2.1. An initial evaluation of HPA-printed microarrays

The first hybridisation using the HPA-printed microarrays was carried out using *F. tularensis* Schu S4 gDNA from the same preparation as that used in the hybridisations described in section 3.3.1. Fewer false positives were observed from the HPA-printed microarray than were observed from the FOI-printed microarray with 0.25 x SSC washes for 15 min. False positive signals that were observed were from negative control probes rather than from buffer or empty spots, indicating that there was little or no print tip carry over (fig. 3.1).

3.3.2.2. Use of salmon DNA to block the microarray from non-specific binding

This was essentially a repeat of the hybridisation described in 3.3.2.1., but with an extra step of blocking the microarray before hybridisation using non-specific salmon testes DNA to reduce the signals from the negative control probes. The hybridisation results were similar to those seen from that described in section 3.3.2.1. (fig. 3.2).

3.3.2.3. Evaluation of an elevated wash temperature to increase stringency

In order to further increase the stringency of the hybridisation, wash steps were carried out at 65°C. Only weak signals were detected from the microarray after this hybridisation (data not shown).

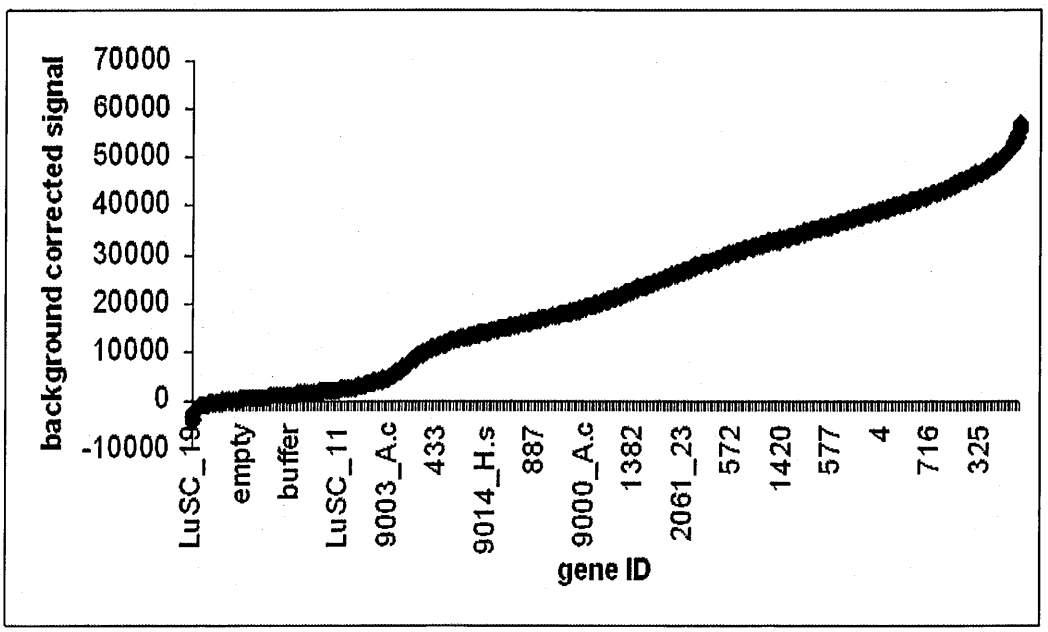


Fig. 3.1 Graph to show the background corrected signal for all probes on the HPA-printed microarray after hybridisation with *F. tularensis* Schu S4 gDNA in DIG Easy Hyb buffer at 42°C. Stringency washes were carried out according to protocol 3. The majority of false positive signals observed were from oligonucleotide probes and not from empty- or buffer only spots, confirming that washing of print pins was adequate.

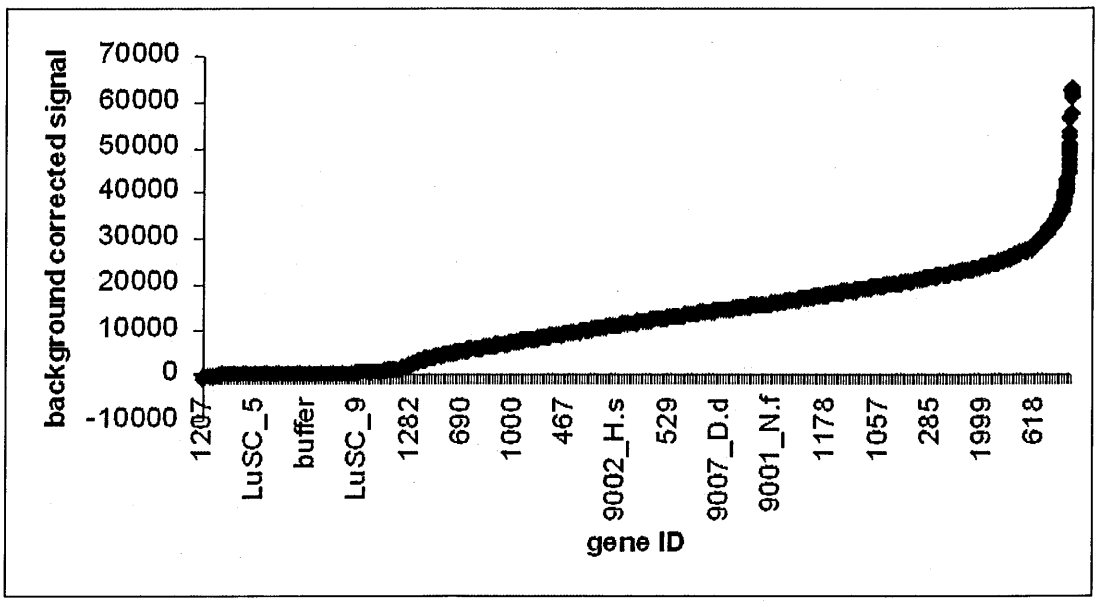


Fig. 3.2 Graph to show the background corrected signal for all probes on the HPA-printed microarray after hybridisation with *F. tularensis* Schu S4 gDNA in DIG Easy Hyb buffer at 42°C. In advance of hybridisation with F. tularensis DNA, the microarray was non-specifically blocked using salmon DNA in an attempt to reduce the signal observed from negative control oligonucleotides. Stringency washes were carried out according to protocol 3. False positive signals were observed from negative control oligonucleotides.

3.3.2.4. A second batch of HPA-printed microarrays

Due to the false positive hybridisations observed, it was thought possible that the oligonucleotides could have become cross contaminated during carriage from FOI to Porton Down, as they were transferred in shallow 96 well microtitre plates. It was not possible to keep the plates upright at all times when travelling (for example to pass the bag through airport security) and although the plates were heat sealed with foil, it was still thought possible that some of the oligonucleotides could have been transferred between wells. This would provide a feasible explanation for the pattern of non-specific hybridisation signals observed from the HPA printed microarrays: some negative control probes, but not all, had high signals, yet the same was not observed from buffer-only or empty control spots. On this basis, a second complete set of oligonucleotides was transferred from FOI to Porton Down for further printing at HPA. Deep 96 well microtitre plates were used, along with an improved heat plate sealing method. Extra care was taken to ensure that the plates were not tipped at all during transit. A second print run of 25 microarrays was carried out at the HPA. This time, more negative control probes were included (32 replicates of each control oligonucleotide), as well as more buffer only and empty spots.

3.3.2.5. Evaluation of the second batch of HPA-printed microarrays

A hybridisation was carried out using a microarray from the second batch of HPA-printed microarrays. The hybridisation was carried out at 42°C in 0.5 x DIG Easy Hyb buffer. Washing was carried out following protocol 3. The results from this first hybridisation showed that only three out of 716

negative control probes give a signal above the background. This is ~0.4% of the total negative control elements and was considered to be an acceptable level of false positive result. The performance of the test probes was also more even than had been observed before (fig. 3.3). However, 103 test probes gave low signals (within the background range) for both replicate spots. This hybridisation was repeated twice more with similar results (figs. 3.4 and 3.5). There were high signals from just 16 (~2.2%) and 11 (~1.5%) of the negative control probes respectively, and none from buffer only or empty spots. It was noted that one negative control probe in particular was responsible for the false positive signal, raising the possibility that there was a level of specific binding between this probe and *F. tularensis* DNA. This oligonucleotide was therefore excluded from further print runs, reducing the false-positive rate to zero.

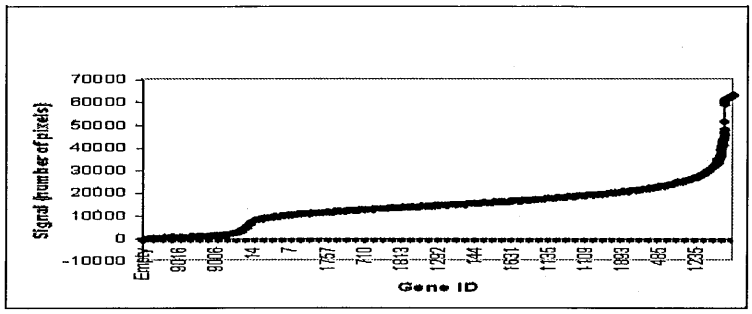


Fig. 3.3 Graph to show the background corrected signal for all probes on the second HPA-printed microarray (using a new set of oligonucleotides) after hybridisation with *F. tularensis* Schu S4 gDNA in DIG Easy Hyb buffer at 42°C. Stringency washes were carried out according to protocol 3. Use of the new set of oligonucleotide probes reduced the false positive rate to 0.4% and produced an even signal from test probes, indicating that previous microarrays had been printed using cross-contaminated oligonucleotides.

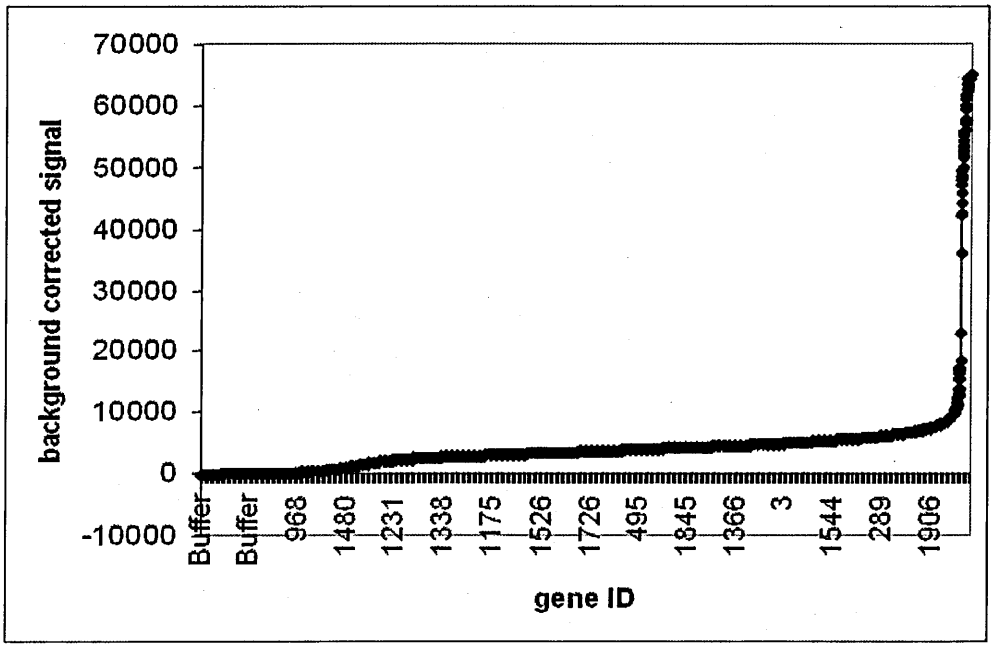


Fig. 3.4 Graph to show the background corrected signal for all probes on the second HPA-printed microarray (using a new set of oligonucleotides) after a second hybridisation with *F. tularensis* Schu S4 gDNA in DIG Easy Hyb buffer at 42°C. Stringency washes were carried out according to protocol 3. The false positive was 2.2% an even signal was observed from test probes, confirming the acceptable result obtained in the previous hybridisation.

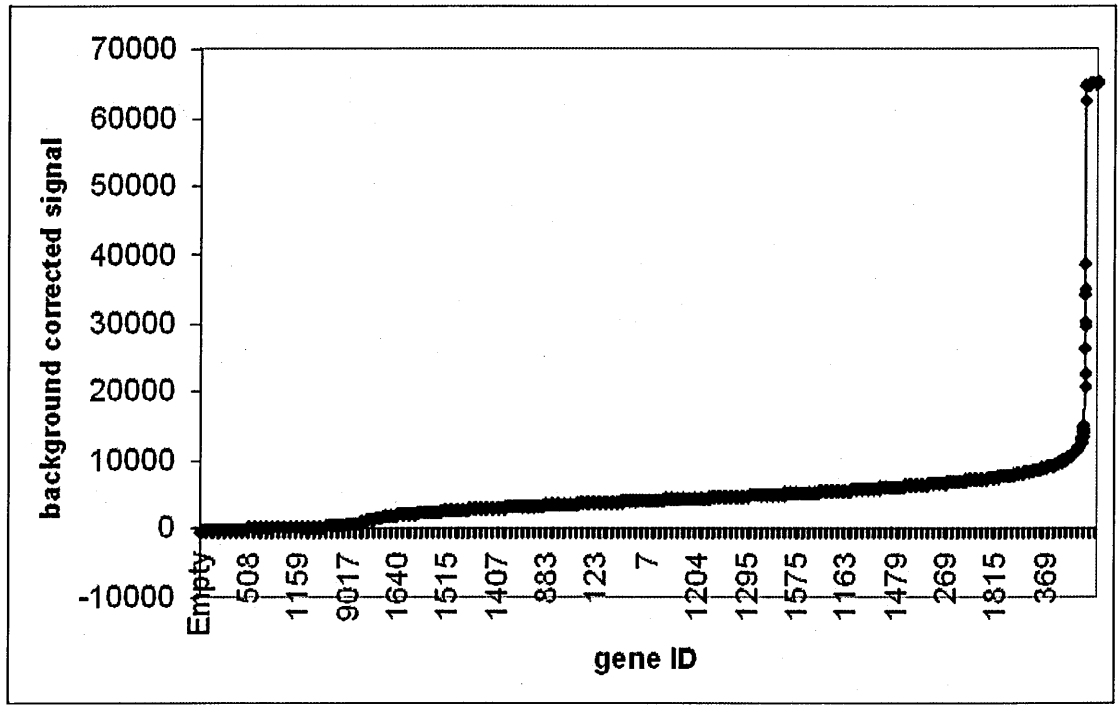


Fig. 3.5 Graph to show the background corrected signal for all probes on the second HPA-printed microarray (using a new set of oligonucleotides) after a third hybridisation with *F. tularensis* Schu S4 gDNA in DIG Easy Hyb buffer at 42°C. Stringency washes were carried out according to protocol 3. The false positive rate was 1.5% and an even signal was observed from test probes.

3.4 Discussion

3.4.1. Oligonucleotide probes were selected for the *F. tularensis* microarray

The choice between oligonucleotide and PCR amplicon probes is usually a trade off between sensitivity and specificity of the probes. Generally, the longer the probe, the more specific its binding capabilities, although such binding specificity can result in reduced sensitivity, particularly, for example, when DNA from a related but non-identical organism is hybridised. Inversely, shorter probes are more likely to share a higher percentage of complementarity with a given target DNA and will therefore be increased in sensitivity, but with a correspondingly lowered specificity. However, in practise the sensitivity of oligonucleotide and PCR amplicon microarrays has been shown to be comparable (Guckenberger *et al.*, 2002). Ideally oligonucleotide probes would be used for applications requiring lowered specificity, such as genotyping, and PCR amplicons would be used to probe samples where higher specificity would be advantageous, such as transcriptomic analysis. An additional advantage of oligonucleotide probes is that these can be chemically synthesised to order, eliminating the necessity of performing, optimising, and purifying the products of a vast number of PCRs and therefore expediting the microarray design process considerably, and it was for this reason that oligonucleotides were selected for the second-generation *F. tularensis* microarray. The length of the oligonucleotide probes was 50 nucleotides, as this length has been reported to provide acceptable levels of both sensitivity and specificity (Kane *et al.*, 2000).

3.4.2. Aminosilane was used to attach the DNA probes to the slides.

A range of chemical coatings can be used to facilitate binding of DNA probes to the solid glass medium, including poly-L-lysine, aldehyde, epoxy-, and amino-silane. Some chemical coatings, such as aldehyde may only be used to facilitate the binding of *amine-modified* DNA to a glass slide. The amino group attached to the DNA binds reversibly with the aldehyde group on the slide to form a Schiff base, which can then be reduced by reaction with sodium borohydride to form an irreversible covalent bond. However, modifying oligonucleotide probes can be costly, and a more economical approach, and one that was taken in the construction of this *F. tularensis* microarray, is to use glass slides that are coated with a chemical that can bind both amine-modified- and un-modified DNA, such as poly-L-lysine, epoxy- or amino-silane, removing the requirement for an additional amine group. Aminosilane and poly-L-lysine both provide a coating of positively-charged amine groups to the slide surface which forms an ionic interaction with the negatively-charged DNA backbone. UV treatment (cross-linking) or baking then causes the conversion of this initial electrostatic interaction into a covalent bond. However, despite the similarities between these two chemicals, it has been reported that retention of DNA onto commercially available poly-L-lysine-coated slides is generally lower than that for commercially available aminosilane-coated slides (Hessner *et al.*, 2004), and for this reason poly-L-lysine coated-slides were not considered for the *F. tularensis* microarray. Epoxy silane works by providing an epoxide group on the slide surface, although a covalent bond between the group and the DNA is formed instantly rather than through cross-linking. At the time of writing,

epoxy-coatings are widely recommended for use with short oligonucleotides, whereas aminosilane is recommended for longer oligonucleotides (> 500-mers) and for PCR products and cDNA (Stef *et al.*, 2006; www.schott.com/nexterion). However, at the time of construction of the second-generation *F. tularensis* microarray, epoxy-coated slides were not widely available, so aminosilane coated slides were chosen as the preferable option.

3.5. Conclusion

The outcome of these experiments was the construction of a second-generation *F. tularensis* Schu S4 microarray that was suitable for use in comparative and functional genomics studies. Through rational optimisation of hybridisation conditions and stringency washes the mean false positive hybridisation rate of the microarray was reduced to zero. The second-generation microarray was based on one defined oligonucleotide probe per CDS annotated in the *F. tularensis* Schu S4 genome, and as such the design was an improvement on those of previously reported microarrays that were constructed using random fragments of *F. tularensis* Schu S4 DNA.

Chapter 4

Comparative genomics of *Francisella*

4.1. Introduction

4.1.1. aCGH

As described in section 1.2.3.3., aCGH is a powerful technique for establishing the similarities and differences between closely related organisms at the DNA level. As mentioned previously, the main advantages of aCGH over other methods are that genome-wide information is generated quickly and relatively inexpensively. The main limitations of aCGH are that all data is generated relative to whichever reference strain(s) are included on the microarray and, unless very high density microarrays are used, small deletions and genome rearrangements are not detectable using this method.

4.1.2. Previous comparative genomics studies of *F. tularensis*

A *F. tularensis* microarray has previously been generated using PCR products derived from the pUC18 clone library synthesised for the genome sequencing of strain Schu S4. The 'shotgun' clones were generated by inserting random DNA fragments into the pUC18 vector. This means that some of the probes contained more than one ORF or truncated ORFs, creating difficulty in assigning addresses to each probe. Samples of gDNA isolated from 27 *F. tularensis* strains were hybridised to this array in single channel reactions to reveal the extent of gross genetic diversity between strains (Broekhuijsen *et al.*, 2003). The results of this study suggest that there is only limited genetic variance between strains within subsp., although eight RD were identified in strains of *F. tularensis* subsp. *holarctica* compared to *F. tularensis* subsp. *tularensis* (Broekhuijsen *et al.*, 2003). A second *F. tularensis* subsp. *tularensis* ordered clone microarray was constructed and

was used to compare the genomes of 16 *F. tularensis* subsp. *tularensis* and subsp. *holarctica* strains that were isolated from the USA, as well as *F. tularensis* LVS under the strain denotation ATCC 29684 (Samrakandi *et al.*, 2004). A high number of polymorphisms observed in this second microarray study suggest that there is significant variability between individual strains. This study identified 13 RD specific to *F. tularensis* subsp. *holarctica*, all associated with repeat sequences or transposases (Samrakandi *et al.*, 2004). A third study of the evolution of the subsp. of *F. tularensis*, sought, in part, to identify RD specific to each *F. tularensis* subsp. (and to *F. novicida*) based on an *in silico* analysis of the position of insertion elements in the *F. tularensis* Schu S4 genome, and confirmed in each strain by PCR (Svensson *et al.*, 2005). This study identified 17 RD between subsp., and the pattern of RD suggested that evolution of *F. tularensis* subsp. occurred through vertical descent, with the most virulent subsp. being the evolutionarily oldest (Svensson *et al.*, 2005). The findings of each of these three studies with respect to RD specific to *F. tularensis* subsp. *holarctica* are summarised in table 4.1.

4.1.3. Aim

The *F. tularensis* microarray (chapter 3) was constructed using the *F. tularensis* Schu S4 genome sequence so that each ORF was represented by an individual oligonucleotide probe, contrasting with and improving upon the ordered clone microarrays used by Broekhuijsen *et al.* (2003) and

Table 4.1. Previously reported *F. tularensis* subsp. *holarctica*-specific RD from each of three studies: Broekhuijsen *et al.*, 2003[a], Samrakandi *et al.*, 2004[b], and Svensson *et al.*, 2005[c]. Some RD were proposed by Samrakandi *et al.* in *F. tularensis* LVS only[d].

	RDs predicted by:		
RD	**Broekhuijsen *et al.*[a]**	**Samrakandi *et al.*[b]**	**Svensson *et al.*, [c]**
1	FTT1580		FTT1580
2	FTT0125 - FTT0126		FTT0125 - FTT0126
3	FTT0960 - FTT0962	FTT0961 - FTT0962	FTT0960 - FTT0962
4	FTT0552 - FTT0553	FTT0552	FTT0552 - FTT0553
5	FTT0521 - FTT0522	FTT0522	FTT0521 - FTT0522
6	FTT1360 – FTT1361	FTT1360 - FTT1362	FTT1360 – FTT1361
7	FTT0426 - FTT0427	FTT0426	FTT0426 - FTT0427
8	FTT1066 - FTT1072	FTT1066 - FTT1073	FTT1066 - FTT1072
9		FTT1308	
10		FTT1581 – FTT1583	
11		FTT0531	FTT0531
13		FTT1789 - FTT1790	
14		FTT1384	
15		FTT0921	
16			FTT1598
17			FTT1484
18 (L3[d])		FTT0918 - FTT0919	FTT1918 – FTT1019
19 (L2[d])		FTT0890	FTT0889 – FTT0890
20			FTT0524
21			FTT0115 – FTT0116
22			FTT0980 – FTT0981
L1[d]		FTT0694	

Samrakandi *et al.* (2004). The aim of this aCGH study was to test the *F. tularensis* Schu S4 microarray by performing hybridisations using gDNA from 12 *Francisella* strains, selected so that at least one strain from each of the *F. tularensis* subsp. was represented, including geographically distinct isolates of *F. tularensis* subsp. *holarctica* and also *F. novicida*. Some of the *F. tularensis* subsp. *holarctica* strains included in this study had also been included in the Broekhuijsen *et al.* (2004) study, allowing for direct comparison of this data with the data generated in that study. It was therefore possible to compare the results of all four *F. tularensis* aCGH studies; ours with those of Broekhuijsen *et al.* (2003), with those of Samrakandi *et al.* (2004), and with those of Svensson *et al.* (2005).

4.2. Methods

4.2.1. Bacterial strains and genomic DNA isolation

Genomic DNA was obtained from each of 12 *F. tularensis* strains of known origin (table 4.2), using the PUREGENE DNA isolation Kit (Gentra Systems), following the manufacturer's protocol, as described in section 2.2.

4.2.2. Microarray hybridization

DNA labelling and hybridization to the *F. tularensis* microarray was carried out as described in section 2.6: For each hybridization 6 μg DNA from the reference strain *F. tularensis* Schu S4 was labelled using Cy5-dCTP and 6 μg DNA from one of the 12 test strains was labelled using Cy3-dCTP. The labelled DNA samples were mixed and competitively hybridized to the array. Hybridizations for each strain were carried out in triplicate.

4.2.3. Data acquisition and analysis

Fluorescent scanning and image quantification was carried out as described in section 2.6.6, using a GenePix 4000B scanner and BlueFuse v.3.1 image analysis software (BlueGnome Ltd., Cambridge, UK).

Spots for which the BlueFuse 'P ON' value (the probability of hybridization) was < 0.5 in the reference channel were excluded from further analysis. Probes were called as differentially hybridizing if the P ON value was < 0.5 in the test channel for both duplicate spots in at least two arrays out of three (Hunter, 2005).

Table 4.2. Details of *Francisella* strains used in this aCGH study

Strain Number	Subsp.	Date	Location	Source	Alternative designation
FSC 237	*tularensis*	1941	Ohio, USA	Human ulcer	Schu S4
FSC 043	*tularensis*	1941	Ohio, USA	Human ulcer	Schu
FSC147	*mediaasiatica*	1965	Kazakhstan	Miday gerbil	GIEM 543
FSC 124	*holarctica*	1990	Ukraine, Odessa region	Water	14588
FSC 257	*holarctica*	1949	Moscow area	Tick	503/840
FSC 012	*holarctica*	1941	Montana, USA	Tick	425F4G
FSC200	*holarctica*	1998	Ljusdal, Sweden	Human	
FSC352	*holarctica*	1974	Nas, Sweden	Hare	
FSC354	*holarctica*	1974	Nas, Sweden	Hare	
FSC358	*holarctica*	1994	Norway	Human	
FSC 338	*holarctica*		Russia	Vaccine strain	Strain 15
FSC 155	*holarctica*		Russia	Vaccine strain	LVS, ATCC 29684
FSC 040	*novicida*	1950	Utah, USA	Water	ATCC 15482

4.2.4. Confirmation of microarray data by PCR

PCRs to confirm microarray data were carried out as described in section 2.3. Primer sequences are given in table 4.3.

4.2.5. Strain- and subsp.-specific RD

RD were identified using the criteria described in section 4.2.3. Previously published RD that have been given in terms of differentiation of *F. tularensis* subsp. *holarctica* (Broekhuijsen *et al.*, 2003; Samrakandi *et al.*, 2004; Svensson *et al.*, 2005) are denoted here as $RD_{holarctica}$. Where possible, $RD_{holarctica}$ are described here in terms of those previously published in order to preserve previous nomenclature. Each $RD_{holarctica}$ may include one RD or multiple RDs if these are adjacent in the *F. tularensis* Schu S4 genome sequence. Similarly, RD identified in this study in *F. tularensis* LVS but not in the progenitor strain *F. tularensis* FSC338 are denoted RD_{LVS}.

Table 4.3. Details of primers used in confirmatory PCR. [a]The predicted amplicon size was based on the *F. tularensis* Schu S4 genome sequence.

Gene	$RD_{holarctica}$	Primer sequence (5'-3')	T_m (°C)	Amplicon size (bp)[a]
oppD	2	F: cgt ggg ttt caa ggc aga tg	54	918
		R: aca aga tgg cac gct gta cg	54	
oppF	2	F: agt agg tga atc cgg ctg tg	54	831
		R: tga caa gct gta taa act tc	46	
FTT1071	8	F: acc tat gcc aga tgc tga ag	52	271
		R: tct tag ata ctc tgg aat tg	46	
FTT0177	23	F: gtc gct aag cga caa gat tc	52	210
		R: ctt ctg ata tag aga aat ag	44	
FTT0446	24	F: gag ttt gct gag cgt tat gg	52	1192
		R: tac cac cca cac cta atg tc	52	
FTT0843	25	F: ggt gct agt tac tgg atg tc	52	1022
		R: agc gca tcg ccg ata att ac	52	
FTT1242	26	F: tca agc aac tgc tgc tca ag	52	901
		R: agg acc agc cac agc act tg	56	
FTT1426	27	F: aac cac aag tcc cag tca ag	52	397
		R: aac cac aag tcc cag tca ag	52	
napH	28	F: tta aga tat tgg tac taa ag	42	317
		R: tgc ctc gcc aga taa tat cg	52	

4.3. Results

Of the 12 *F. tularensis* strains selected for aCGH (table 4.2), nine were *F. tularensis* subsp. *holarctica* from varying geographical regions including *F. tularensis* LVS, one was *F. tularensis* subsp. *mediaasiatica* and one was *F. tularensis* subsp. *tularensis*. The remaining sample was an *F. novicida* strain. As a control, the reference strain *F. tularensis* Schu S4 was hybridized in the same way as a test strain, so that these positive control microarrays were hybridized with strain Schu S4 DNA in both channels.

4.3.1. RD identified by aCGH compared to those predicted by genome sequence.

The genome sequence of FSC155, *F. tularensis* LVS, has been determined by The Lawrence Livermore National Laboratory, California, USA (Chain *et al.*, 2006), allowing an *in silico* comparison with the genome sequence of the reference strain *F. tularensis* Schu S4 the tBLASTx (Altschul *et al.*, 1990) to compare translated nucleotide ORFs with translated nucleotide sequences in the query strain. A comparison of the RD identified in this strain by aCGH with the ORFs predicted to be absent from *F. tularensis* LVS by genome sequence comparison was then made (fig. 4.1.). Of the 70 RD detected by aCGH, only 20 were also predicted by comparison of the genome sequences. There were 13 RD predicted by comparison of the genome sequences that were not detected by aCGH. Nine of the RD predicted by aCGH and not by sequence comparison coincided with previously published $RD_{holarctica}$, whereas only one RD predicted by sequence

ID	FTT no.	RD[c]	Array[a]	Seq.[b]
sdhC	FTT0072			
acnA	FTT0087			
FTT0095	FTT0095			
FTT0112	FTT0112			
prfC	FTT0118			
oppA	FTT0122			
oppF	FTT0126	2		
nusG	FTT0139			
FTT0156	FTT0156			
FTT0177	FTT0177	23		
ftsA	FTT0187			
purH	FTT0203			
usp	FTT0245			
FTT0382	FTT0382			
glgC	FTT0415			
glyQ	FTT0419			
FTT0426	*thrA*	7		
ostA1	FTT0467			
FTT0488	FTT0488			
FTT0522	FTT0522	5		
dinP	FTT0529			
FTT0552	FTT0552	4		
cynT	FTT0592			
hfq	FTT0630			
ilvD	FTT0640			
grxB	FTT0650			
lipA	FTT0653			
FTT0719	FTT0719			
FTT0744	FTT0744			
FTT0747	FTT0747			
FTT0843	FTT0843	25		
FTT0844	FTT0844	25		
FTT0845	FTT0845	25		
FTT0851	FTT0851			
FTT0889	FTT0889			
FTT0919	FTT0919	18		
FTT0921	FTT0921	15		
folK	FTT0942			
FTT0960	FTT0960	3		
mdaB	FTT0961	3		
FTT0962	FTT0962	3		

ID	FTT no.	RD[c]	Array[a]	Seq.[b]
aroG	FTT0963			
ffh	FTT0964			
FTT0965	FTT0965			
FTT0982	FTT0982			
FTT0983	FTT0983			
pheS	FTT1003			
rimI	FTT1054			
FTT1066	FTT1066	8		
res	FTT1067	8		
FTT1068	FTT1068	8		
FTT1069	FTT1069	8		
FTT1070	FTT1070	8		
FTT1071	FTT1071	8		
FTT1072	FTT1072	8		
FTT1073	FTT1073	8		
alaS	FTT1096			
FTT1193	FTT1193			
dedA2	FTT1222			
FTT1242	FTT1242	26		
FTT1252	FTT1252			
yhhW	FTT1266			
FTT1308	FTT1308	9		
pdpD	FTT1360	6		
FTT1361	FTT1361	6		
panC	FTT1390			
FTT1423	FTT1423			
FTT1443	FTT1443			
ppx	FTT1444			
wbtJ	FTT1454			
gyrA	FTT1575			
FTT1579	FTT1579			
FTT1580	FTT1580	1		
FTT1582	FTT1582	10		
FTT1583	FTT1583	10		
napH	FTT1670	28		
FTT1716	FTT1716			
FTT1719	FTT1719			
kdpC	FTT1737			
recX	FTT1751			
FTT1790	FTT1790			
FTT1791	FTT1791	13		
16s rRNA				

Fig. 4.1 A comparison of RD predicted in *F. tularensis* LVS by aCGH[a] and those predicted by comparison of the *F. tularensis* Schu S4 and LVS genome sequences[b], according to the key below. Coincidence of RD identified in *F. tularensis* LVS (by either genome sequence comparison or by aCGH) with $RD_{holarctica}$ are also shown[c].

Absent Marginal Present Microarray and sequence comparison agreement

comparison but not by aCGH coincided with a previously published $RD_{holarctica}$.

4.3.2. Confirmation of results by PCR

In order to validate the microarray results, regions of eight ORFs from $RD_{holarctica}$ including three that were previously published and both supported and not supported by the data from this study were amplified by PCR. PCR primers were designed so that amplicons would include the complementary sequences of the corresponding microarray oligonucleotide probes. The results of seven out of eight PCR reactions fully supported the microarray data observed in this study (fig. 4.2. and table 4.5.). The reaction to amplify FTT1426 ($RD_{holarctica}27$ [section 4.3.5.16.]) only partially confirmed the microarray findings in that amplicons of the correct size were observed from the following *F. tularensis* subsp. *holarctica* templates: FSC200, FSC124, and FSC257, meaning that this RD could not be described as a $RD_{holarctica}$.

4.3.3. Hybridisations using DNA from *F. tularensis* Schu S4

Control hybridisations were carried out in triplicate using DNA from *F. tularensis* Schu S4, which was the strain to which the microarray was designed (chapter 3). No RD were observed in the reference strain *F. tularensis* Schu S4.

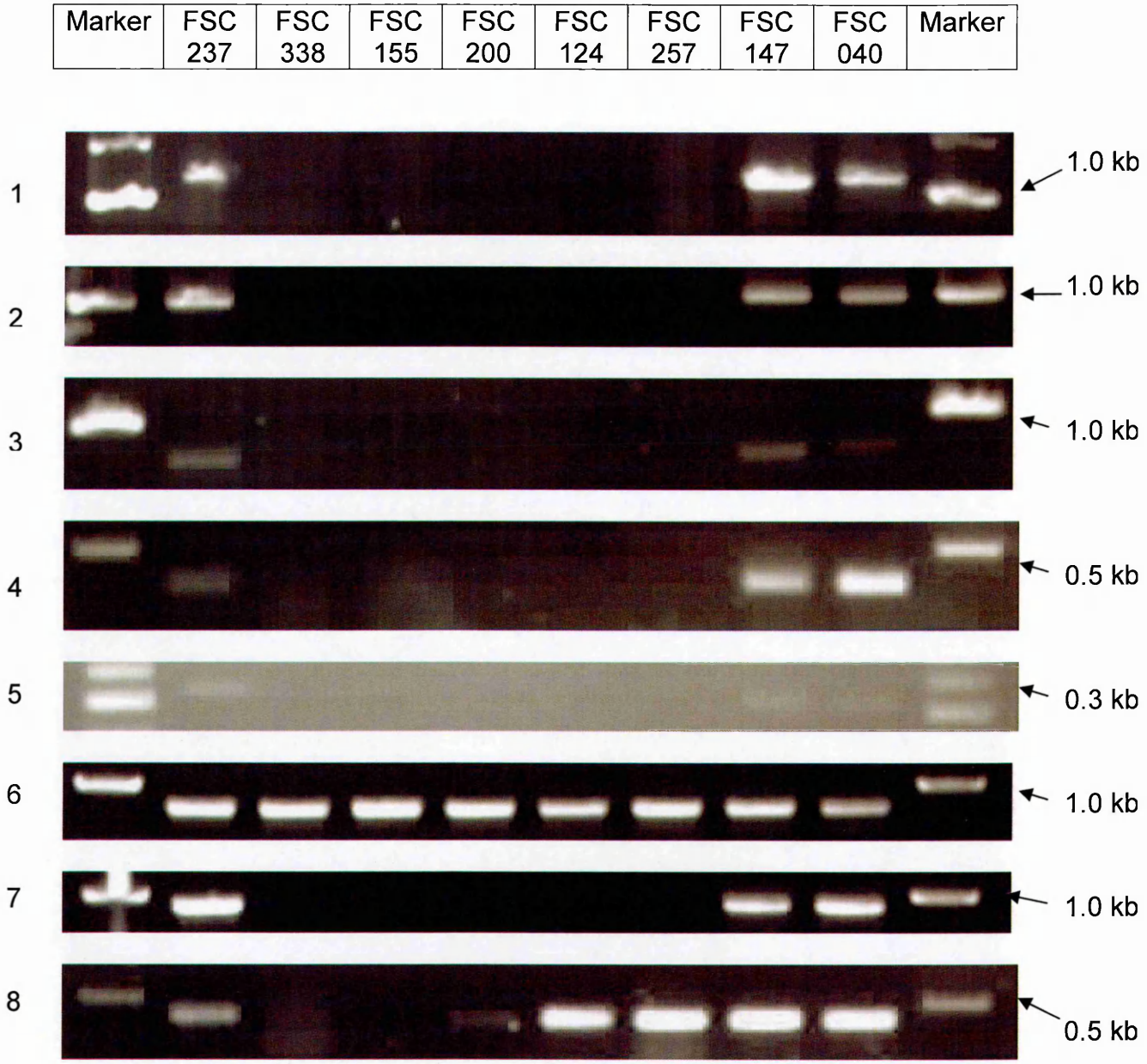


Fig. 4.2 PCRs were used to confirm aCGH data by amplification of regions of eight ORFs from $RD_{holarctica}$ that were both supported and not supported by the data obtained in this aCGH study, according to the key in table 4.4. The results of seven out of eight PCR reactions fully supported the microarray data observed in this study. However, when using primers complimentary to FTT1426 ($RD_{holarctica}27$), PCR amplicons of the correct size were observed from three *F. tularensis* subsp. *holarctica* strains, meaning that this RD could not be described as a $RD_{holarctica}$.

Table 4.4 Key to fig. 4.2

Gel	ORF	$RD_{holarctica}$	Expected amplicon size (bp)[a]
1	FTT0446	24	1192
2	FTT0843	25	1022
3	FTT1242	26	901
4	*napH*	29	317
5	FTT1071	8	271
6	*oppD*	2	918
7	*oppF*	2	831
8	FTT1426	28	397

[a]based on the *F. tularensis* Schu S4 genome sequence

Table 4.5 Confirmatory PCR results. ORFs for which an amplicon was obtained by PCR at the expected size (based on the *F. tularensis* Schu S4 genome sequence) are denoted positive (+) here, and ORFs for which no amplicon was obtained by PCR are denoted negative (-) here. The colour of each square indicates the result for each ORF in the specific strain by aCGH according to the key below.

Present by aCGH
Marginal by aCGH
Absent by aCGH
No aCGH data

	Gene	*oppD*	*oppF*	FTT1071	FTT0446	FTT0843	FTT1242	FTT1426	*napH*
	$RD_{holarctica}$	2	2	8	24	25	26	27	28
FSC237		+	+	+	+	+	+	+	+
FSC338		+	-	-	-	-	-	-	-
FSC155		+	-	-	-	-	-	-	-
FSC200		+	-	-	-	-	-	+	-
FSC124		+	-	-	-	-	-	+	-
FSC257		+	-	-	-	-	-	+	-
FSC147		+	+	+	+	+	+	+	+
FSC040		+	+	+	+	+	+	+	+

4.3.4. Genomic differences identified using the *F. tularensis* microarray

Fig. 4.3 is a heat map comparing RD patterns for all strains tested. RDs totalled between 0.05% and 17.32% of the total number of probes, depending on the particular strain.

4.3.4.1. RD identified in *F. novicida*

DNA from *F. novicida*, FSC040, showed a hybridization pattern that was distinct from those of any of the other strains tested. However, only seven RD were identified (0.47% of the probes): *capB*, *capC*, *gcvP1*, *leuS*, *lpnB*, *ribD*, *udk*, oligo1668, and oligo1670. None of these RD were coincident with any $RD_{holarctica}$.

4.3.4.2. RD identified in *F. tularensis* subsp. *tularensis*

No RD were observed for the microarray reference strain *F. tularensis* Schu S4. The avirulent *F. tularensis* subsp. *tularensis* strain FSC043 showed a high degree of genetic similarity to the reference strain Schu S4, with only one RD identified, FTT0918, which was coincident with $RD_{holarctica}18$. In the *F. tularensis* Schu S4 genome sequence FTT0918 is annotated as a hypothetical protein.

4.3.4.3. RD identified in *F. tularensis* subsp. *mediasiatica*

The *F. tularensis* subsp. *mediaasiatica* strain tested here, FSC147, showed a high degree of genetic similarity with the reference strain Schu S4, with only three RD identified in this study: FTT0176, FTT1362, and FTT1733. Two of these RD were coincident with $RD_{holarctica}$.

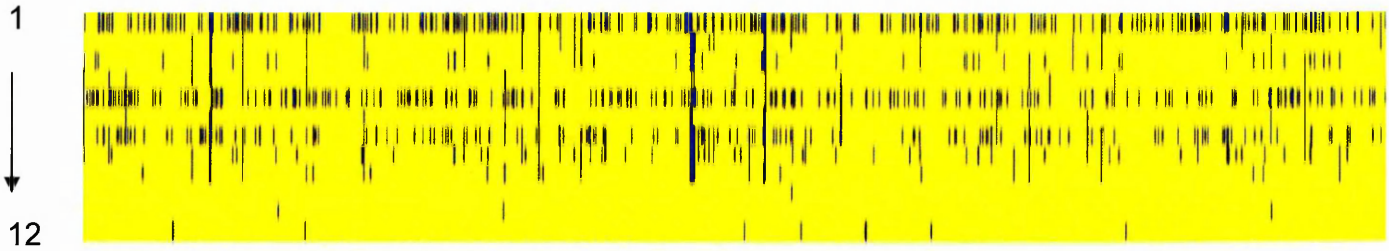


Fig. 4.3 A 'heat map' showing aCGH data observed for each probe in all *Francisella* strains tested. Columns in the horizontal axis represent each oligonucleotide included on the *F. tularensis* Schu S4 microarray and the vertical axis represents *Francisella* strains tested according to the table below. Blue bars represent probes which were absent or divergent in by aCGH, and yellow bars represent probes which were present by aCGH.

Column	Strain
1	*F. tularensis* subsp. *holarctica* FSC012
2	*F. tularensis* subsp. *holarctica* FSC338
3	*F. tularensis* subsp. *holarctica* FSC155
4	*F. tularensis* subsp. *holarctica* FSC200
5	*F. tularensis* subsp. *holarctica* FSC352
6	*F. tularensis* subsp. *holarctica* FSC354
7	*F. tularensis* subsp. *holarctica* FSC358
8	*F. tularensis* subsp. *holarctica* FSC124
9	*F. tularensis* subsp. *holarctica* FSC257
10	*F. tularensis* subsp. *tularensis* FSC043
11	*F. tularensis* subsp. *mediaasiatica* FSC147
12	*F. novicida* FSC040

4.3.4.4. RD identified in *F. tularensis* subsp. *holarctica*

Overall it was *F. tularensis* subsp. *holarctica* that displayed the highest level of diversity by microarray to the reference strain *F. tularensis* Schu S4 (figs. 4.4. to 4.12). The most divergent strain was FSC012, which was the only *F. tularensis* subsp. *holarctica* strain from the USA that was tested. Many of the RDs were unique to each strain tested although only one was entirely unique to *F. tularensis* LVS (FSC155). The RD unique to FSC155 was oligo766, a probe designed to an intergenic region that was designated an ORF during the first GLIMMER analysis of the (incomplete) genome sequence, but designated intergenic in the final annotation of the completed genome sequence. In *F. tularensis* Schu S4 Oligo766, maps to the intergenic region between FTT0492 (*lysR*) and FTT0494 (*cutC*).

4.3.5. $RD_{holarctica}$

As described in 4.2.5. $RD_{holarctica}$ were common to all or most *F. tularensis* subsp. *holarctica* strains compared to the other three *F. tularensis* subsp. tested. The data obtained from *F. tularensis* subsp. *holarctica* in this study were examined in the context of previously published $RD_{holarctica}$ to investigate the level of similarity between these results and those previously published. Based on $RD_{holarctica}$ identified in this study, 17 $RD_{holarctica}$ are proposed, eleven of which have been previously proposed by Broekhuijsen *et al.* (2003) and/or Samrakandi *et al.* (2004) and/or Svensson *et al.* (2005) (fig. 4.13, table 4.7). $RD_{holarctica}$ that have been previously proposed have been designated with the same $RD_{holarctica}$ number here.

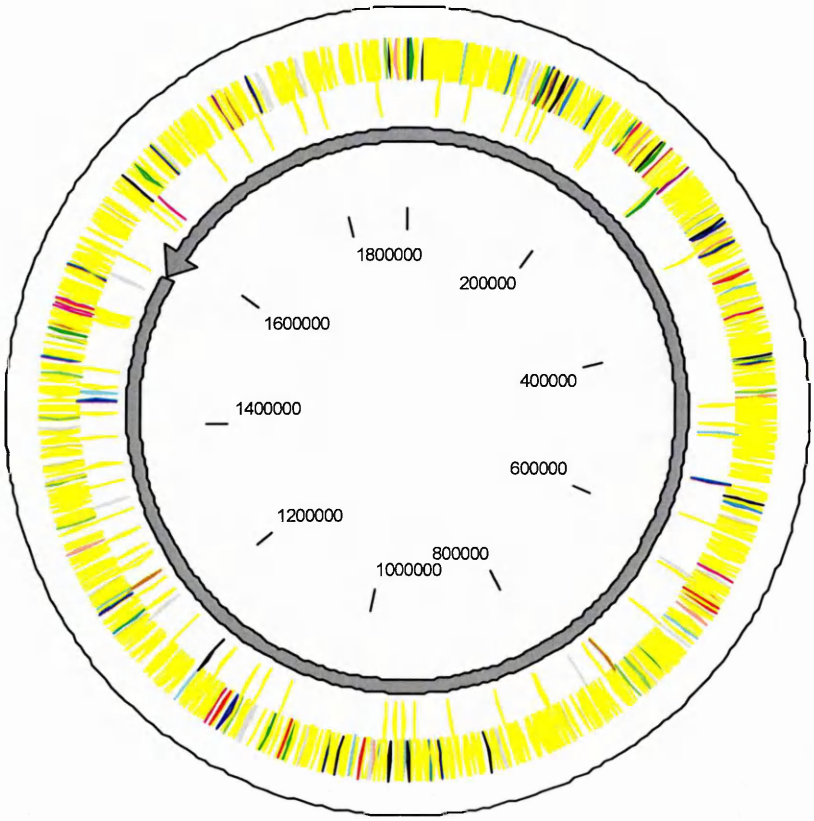


Fig.4.4 RD observed in *F. tularensis* subsp. *holarctica*, FSC012. FSC012 was isolated in the USA. Genes present in FSC012 are shown here in yellow, DHPs are indicated by bars coloured according to their broad function or orthologous group (table 4.6). FSC012 showed 17.32% divergence from the reference strain *F. tularensis* subsp. *tularensis* strain Schu S4 by aCGH.

Table 4.6 Colouring of orthologous groups used in figs. 4.4 - 4.12

Energy production and conversion	
Cell division and chromosome partitioning	
Amino acid transport and metabolism	
Nucleotide transport and metabolism	
Carbohydrate transport and metabolism	
Coenzyme metabolism	
Lipid metabolism	
Transcription and translation	
DNA replication, recombination, and repair	
Cell envelope biogenesis, outer membrane	
Post translational modification, protein turnover, chaperones	
Inorganic ion transport	
Signal transduction mechanisms	
Unknown function/general function prediction only	
Pseudogene	

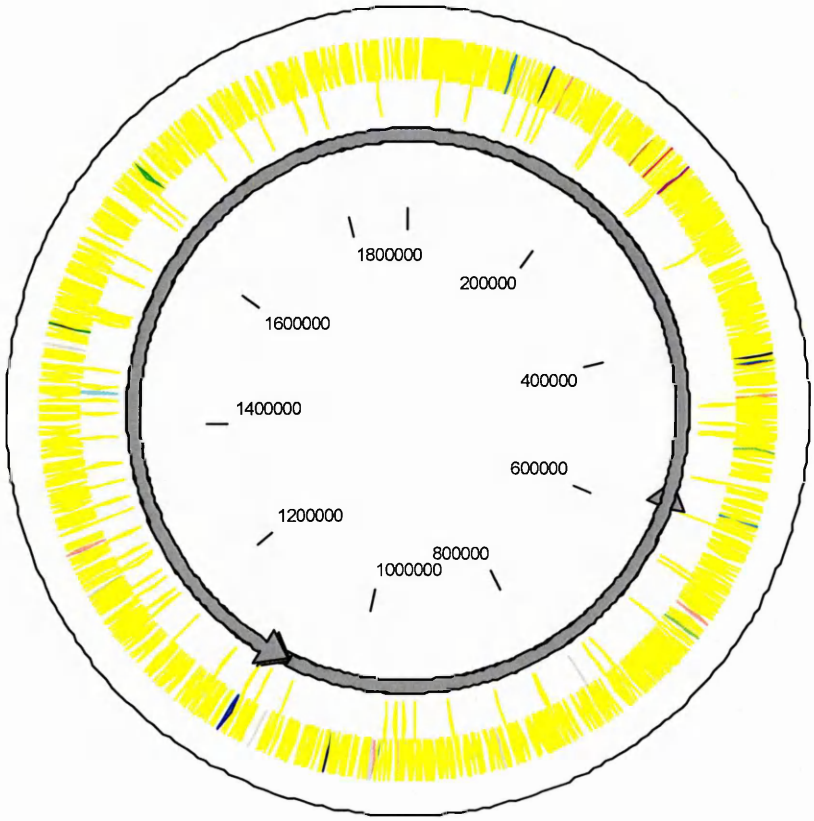


Fig. 4.5 RD observed in *F. tularensis* subsp. *holarctica*, FSC338. FSC338 is a Russian laboratory-derived vaccine strain, thought to be the parental strain of FSC155 (Tigertt, 1962). Genes present in FSC338 are shown here in yellow, DHPs are indicated by bars coloured according to their broad function or orthologous group (table 4.6). FSC338 showed 3.18% divergence from the reference strain *F. tularensis* subsp. *tularensis* strain Schu S4 aCGH.

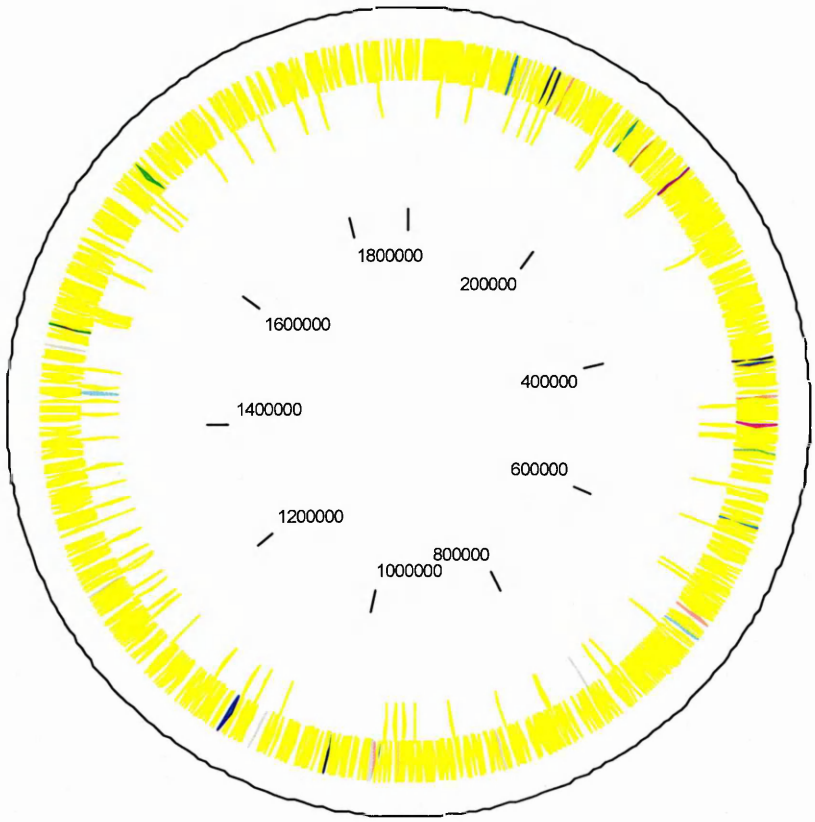


Fig. 4.6 RD observed in *F. tularensis* subsp. *holarctica*, FSC155. This is *F. tularensis* LVS, a Russian laboratory-derived strain. Genes present in FSC155 are shown here in yellow, DHPs are indicated by bars coloured according to their broad function or orthologous group (table 4.6). FSC155 showed 3.54% divergence from the reference strain *F. tularensis* subsp. *tularensis* strain Schu S4 aCGH.

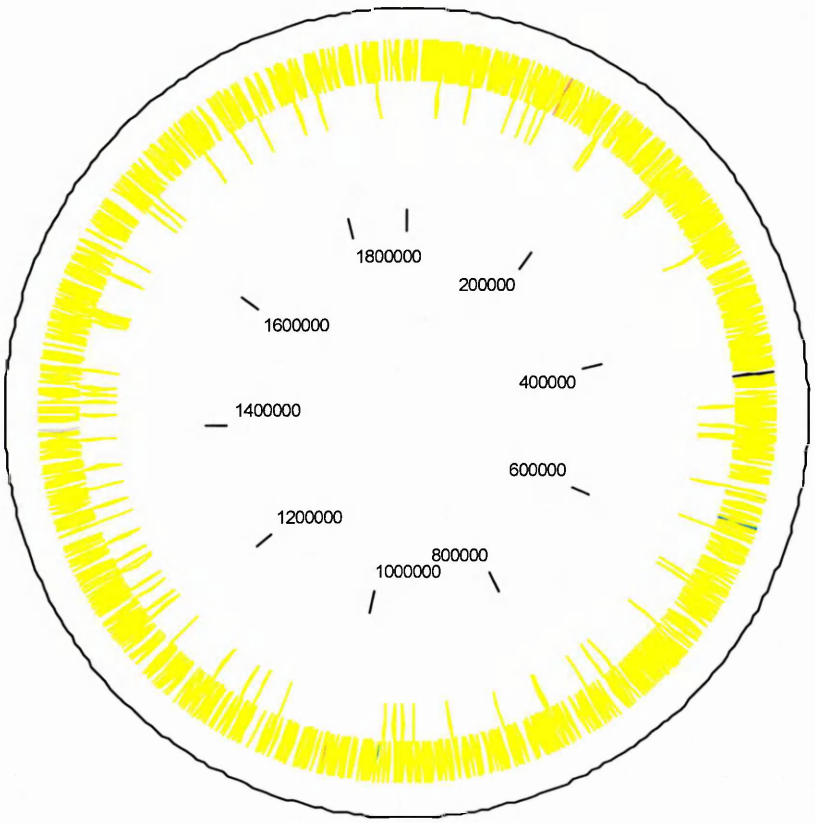


Fig. 4.7 RD observed in *F. tularensis* subsp. *holarctica*, FSC200. FSC200 was isolated in Sweden. Genes present in FSC200 are shown here in yellow, DHPs are indicated by bars coloured according to their broad function or orthologous group (table 4.6). FSC200 showed 1.09% divergence from the reference strain *F. tularensis* subsp. *tularensis* strain Schu S4 aCGH.

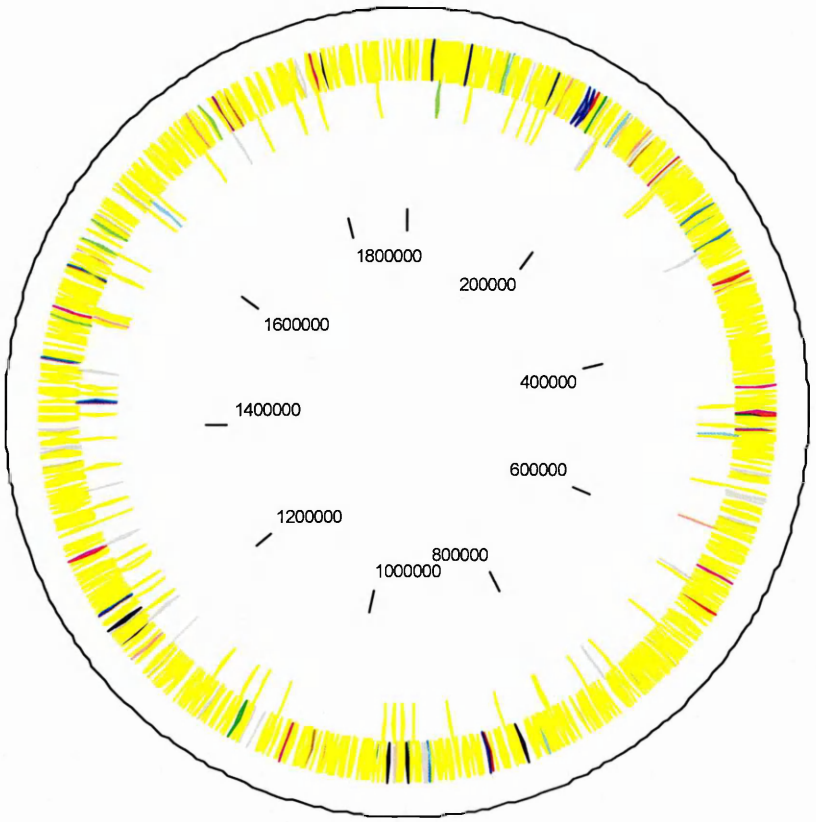


Fig. 4.8 RD observed in *F. tularensis* subsp. *holarctica*, FSC352. FSC352 is a Swedish isolate. Genes present in FSC352 are shown here in yellow, DHPs are indicated by bars coloured according to their broad function or orthologous group (table 4.6). FSC352 showed 41.09% divergence from the reference strain *F. tularensis* subsp. *tularensis* strain Schu S4 aCGH.

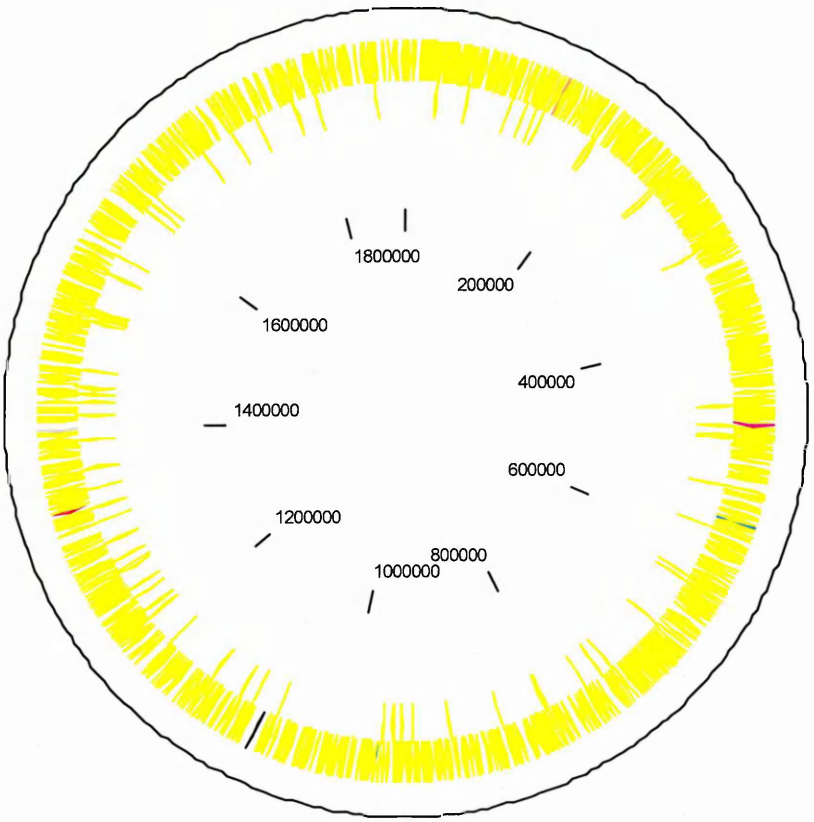


Fig. 4.9 RD observed in *F. tularensis* subsp. *holarctica*, FSC354. FSC354 was isolated in Sweden. Genes present in FSC354 are shown here in yellow, DHPs are indicated by bars coloured according to their broad function or orthologous group (table 4.6). FSC354 showed 1.09% divergence from the reference strain *F. tularensis* subsp. *tularensis* strain Schu S4 aCGH.

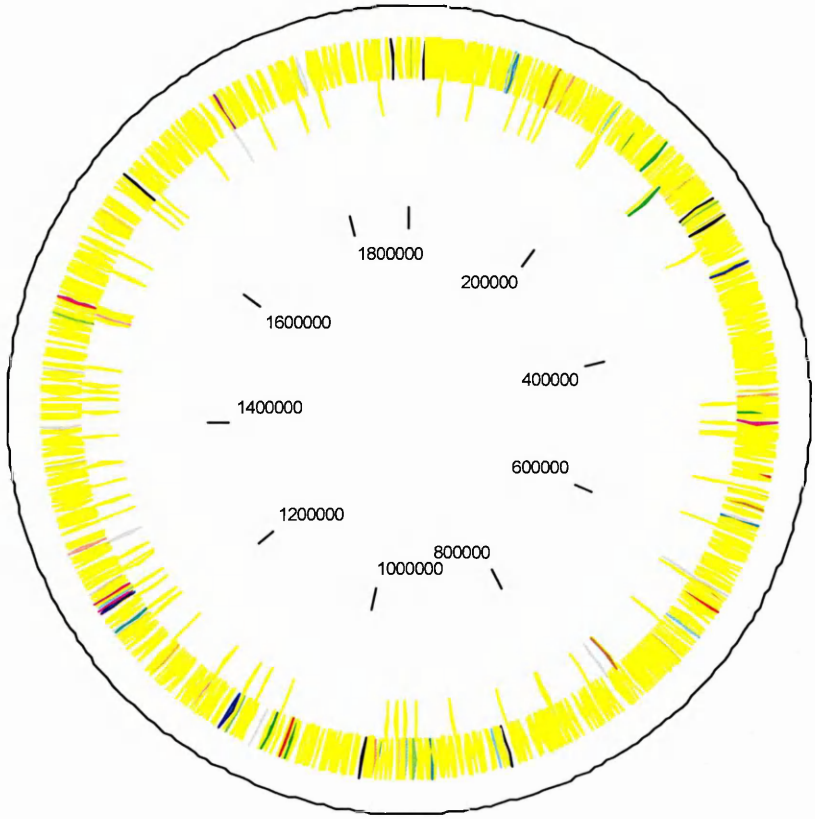


Fig. 4.10 RD observed in *F. tularensis* subsp. *holarctica*, FSC358. FSC358 was isolated in Norway. Genes present in FSC358 are shown here in yellow, DHPs are indicated by bars coloured according to their broad function or orthologous group (table 4.6). FSC358 showed 9.62% divergence from the reference strain *F. tularensis* subsp. *tularensis* strain Schu S4 by aCGH.

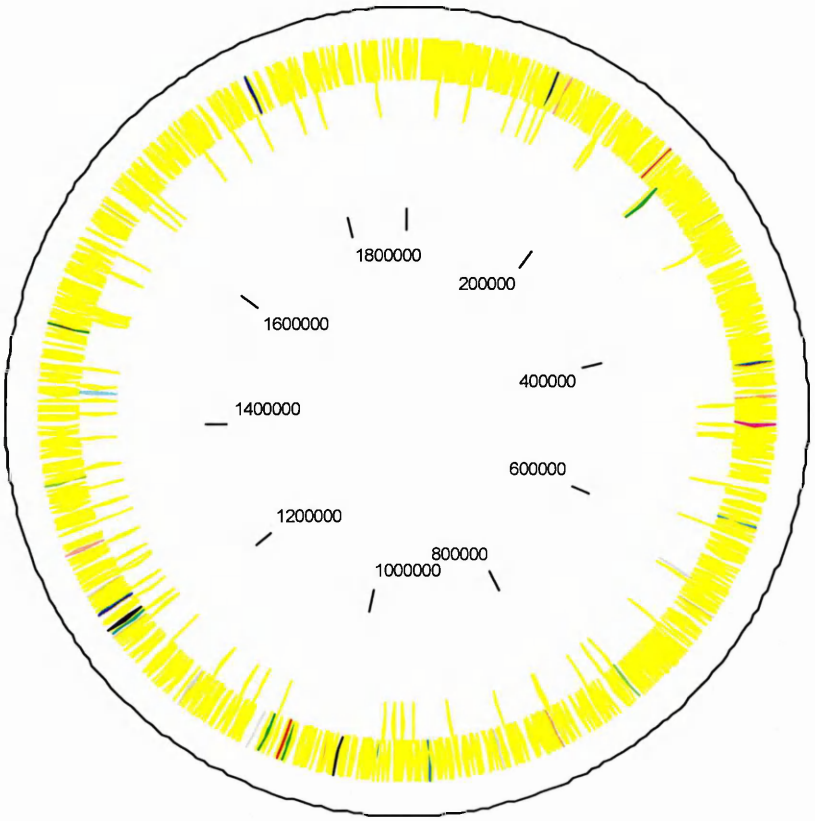


Fig. 4.11 RD observed in *F. tularensis* subsp. *holarctica*, FSC124. FSC124 was isolated in the Ukraine. Genes present in FSC124 are shown here in yellow, DHPs are indicated by bars coloured according to their broad function or orthologous group (table 4.6). FSC124 showed 4.21% divergence from the reference strain *F. tularensis* subsp. *tularensis* strain Schu S4 aCGH.

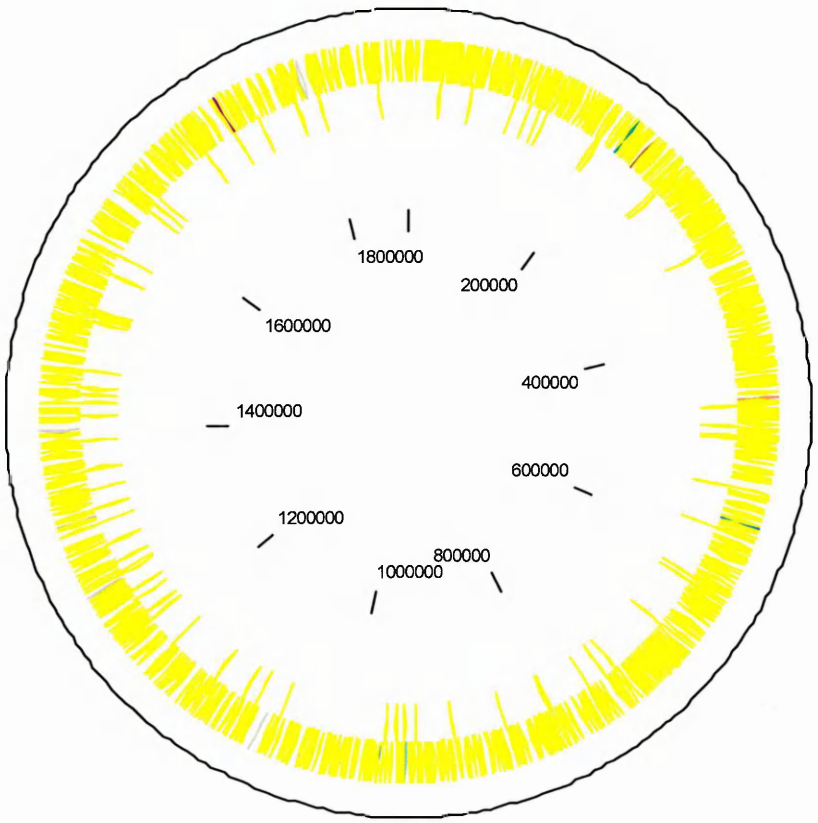


Fig. 4.12 RD observed in *F. tularensis* subsp. *holarctica*, FSC257. SC257 was isolated in the Moscow. Genes present in FSC257 are shown here in yellow, DHPs are indicated by bars coloured according to their broad function or orthologous group (table 4.6). FSC257 showed 1.66% divergence from the reference strain *F. tularensis* subsp. *tularensis* strain Schu S4 by aCGH.

Figure 4.13 $RD_{holarctica}$ data obtained by aCGH for each *Francisella* strain tested

ORF present by aCGH

ORF absent or divergent by aCGH

ORF marginal: within 10% of absent cut off

No data obtained

[a]$RD_{holarctica}$ previously published but not supported by data from this study

[b]Described as L3 by Samrakandi *et al.*, (2004)

[c]Described as L2 by Samrakandi *et al.*, (2003)

[d]Pseudogene in *F. tularensis* Schu S4

$RD_{holarctica}$ previously proposed by:

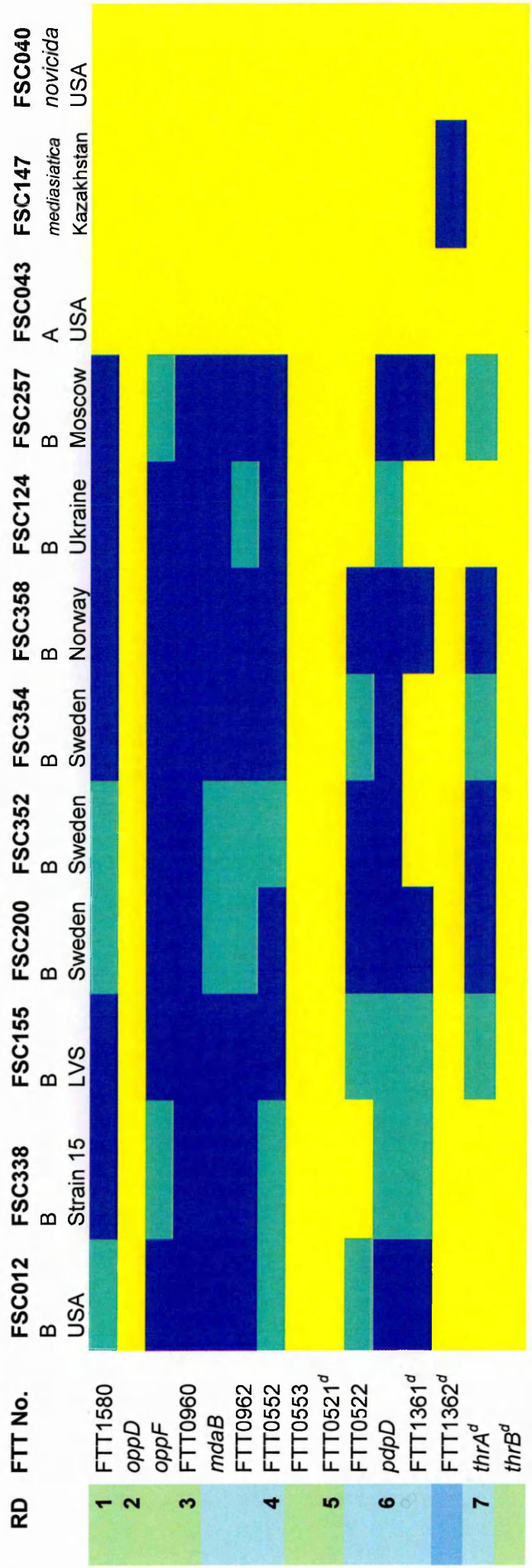
Broekhuijsen *et al.*, 2003 and Svensson *et al.*, 2005

Broekhuijsen *et al.*, 2003, Samrakandi *et al.*, 2004, and Svensson *et al.*, 2005

Samrakandi *et al.*, 2004

Svensson *et al.*, 2005

Samrakandi *et al.*, 2004 and Svensson *et al.*, 2005


RD
FTT No.
FSC012 B USA
FSC338 B Strain 15
FSC155 B LVS
FSC200 B Sweden
FSC352 B Sweden
FSC354 B Sweden
FSC358 B Norway
FSC124 B Ukraine
FSC257 B Moscow
FSC043 A USA
FSC147 mediasiatica Kazakhstan
FSC040 novicida USA
1
2
3
4
5
6
7
FTT1580
oppD
oppF
FTT0960
mdaB
FTT0962
FTT0552
FTT0553
FTT0521d
FTT0522
pdpD
FTT1361d
FTT1362d
thrAd
thrBd

RD
FTT No.
FSC012 B USA
FSC338 B Strain 15
FSC155 B LVS
FSC200 B Sweden
FSC352 B Sweden
FSC354 B Sweden
FSC358 B Norway
FSC124 B Ukraine
FSC257 B Moscow
FSC043 A USA
FSC147 mediasiatica Kazakhstan
FSC040 novicida USA
8
FTT1066
resd
FTT1068
FTT1069
FTT1070d
FTT1071
FTT1072
FTT1073
FTT1075
hipA
FTT1077
FTT1078
FTT1079
FTT1080
FTT1082
FTT1083
9
10
FTT1308
FTT1581
FTT1582d
FTT1583d
11a
13
FTT0531
FTT1789
FTT1790d
FTT1791
14a
15
FTT1384
FTT0921d
16a
FTT1598

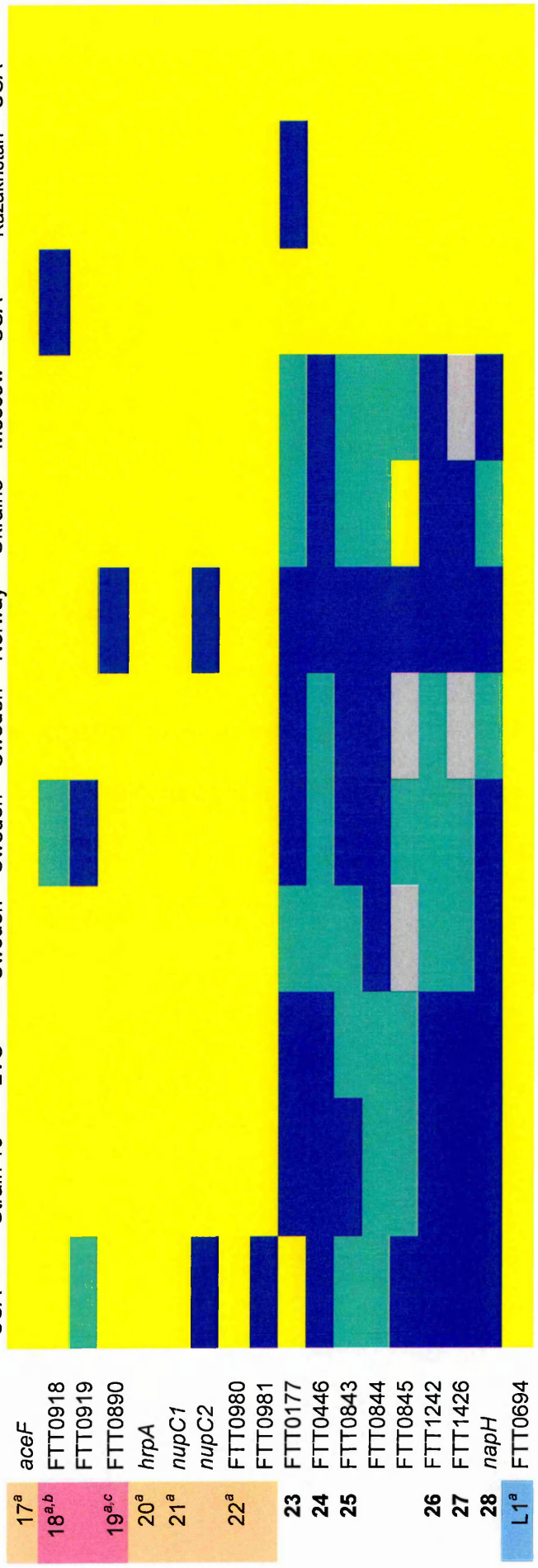

RD
FTT No.
FSC012 B USA
FSC338 B Strain 15
FSC155 B LVS
FSC200 B Sweden
FSC352 B Sweden
FSC354 B Sweden
FSC358 B Norway
FSC124 B Ukraine
FSC257 B Moscow
FSC043 A USA
FSC147 mediasiatica Kazakhstan
FSC040 novicida USA
17a
aceF
18a,b
FTT0918
FTT0919
19a,c
FTT0890
20a
hrpA
21a
nupC1
nupC2
22a
FTT0980
FTT0981
23
FTT0177
24
FTT0446
25
FTT0843
FTT0844
FTT0845
26
FTT1242
27
FTT1426
28
napH
L1a
FTT0694

Table 4.7 $RD_{holarctica}$ predicted from this study and $RD_{holarctica}$ previously reported by Broekhuijsen *et al.*, 2003[a], Samrakandi *et al.*, 2004[b], and Svensson *et al.*, 2005[c]. Some RD were proposed by Samrakandi *et al.* in *F. tularensis* LVS only[d]. Data obtained from this study indicated that $RD_{holarctica}6$ could not be confirmed for all strains tested[e].

	RDs predicted by:			
RD	**Broekhuijsen *et al.*[a]**	**Samrakandi *et al.*[b]**	**Svensson *et al.*, [c]**	**This study**
1	FTT1580 - FTT1581		FTT1580 - FTT1581	FTT1580
2	FTT0125 - FTT0126		FTT0125 - FTT0126	FTT0126
3	FTT0960 - FTT0962	FTT0961 - FTT0962	FTT0960 - FTT0962	FTT0960 - FTT0962
4	FTT0552 - FTT0553	FTT0552	FTT0552 - FTT0553	FTT0552
5	FTT0521 - FTT0522	FTT0522	FTT0521 - FTT0522	FTT0522 (partial)
6	FTT1360 - FTT1361	FTT1360 - FTT1362	FTT1360 - FTT1361	FTT1360 - FTT1361[e]
7	FTT0426 - FTT0427	FTT0426	FTT0426 - FTT0427	FTT0426 (partial)
8	FTT1066 - FTT1072	FTT1066 - FTT1073	FTT1066 - FTT1072	FTT1066 - FTT1073 (USA strain FTT1071-83)
9		FTT1308		FTT1308
10		FTT1581 - FTT1583		FTT1582 - FTT1583
11		FTT0531	FTT0531	
13		FTT1789 - FTT1790		FTT1791 (partial)
14		FTT1384		
15		FTT0921		FTT0921
16			FTT1598	
17			FTT1484	
18 (L3[d])		FTT0918 - FTT0919	FTT0918 - FTT0919	
19 (L2[d])		FTT0890	FTT0889 - FTT0890	
20			FTT0524	
21			FTT0115 - FTT0116	
22			FTT0980 - FTT0981	
23				FTT0177
24				FTT0446
25				FTT0843 – FTT0845
26				FTT1242
27				FTT1426
28				FTT1670
L1[d]		FTT0694		

4.3.5.1. RD*holarctica*1

The RD*holarctica*1 proposed by Broekhuijsen *et al.* was a variable chromosomal region consisting of two ORFs (four in *F. tularensis* subsp. *novicida*) which map to ORFs FTT1580 and FTT1581. Investigation of this region by Broekhuijsen *et al.* showed that 1) both ORFs were present in *F. tularensis* subsp. *tularensis* strains; 2) both ORFs were present in *F. tularensis* subsp. *mediasiatica* (with FTT1580 truncated by 10 amino acids from the C terminus and FTT1581 truncated by 12 amino acids from the N terminus); 3) FTT1580 was absent in European *F. tularensis* subsp. *holarctica* strains; 4) FTT1581 was absent in *F. tularensis* subsp. *holarctica* biovar *japonica* strains; and 5) in *F. novicida* this region was composed of four ORFs. The microarray results observed in this study support the findings of Broekhuijsen *et al.* in that both FTT1580 and FTT1581 were present for the *F. tularensis* subsp. *tularensis*, subsp. *mediasiatica* and subsp. *novicida* strains tested but FTT1580 was absent or marginal (within 10% of the cut off criteria for all probes not excluded) for the *F. tularensis* subsp. *holarctica* strains tested.

4.3.5.2. RD*holarctica*2

The RD*holarctica*2 proposed by Broekhuijsen *et al.* includes the ORFs FTT0125 and FTT0126, *oppD* and *oppF,* respectively. By aCGH *oppF* was absent or marginal in all subsp. *holarctica* strains tested. The ORF *oppD* was present in all strains tested.

4.3.5.3. RD*holarctica*3

Both Broekhuijsen *et al.* and Samrakandi *et al.* proposed RD*holarctica* 3: three ORFs were predicted by Broekhuijsen *et al.*: *mdaB*, a conserved hypothetical protein and a hypothetical arginine-orthinine operon transport protein. Samrakandi *et al.* predicted two ORFs: *mdaB* and a putative ABC transporter. The results from this study show RD3 to consist of three ORFs: *mdaB*, FTT0961 (conserved hypothetical protein), and *xthA* (annotated as exodeoxyribonuclease III). Although the annotations vary between each of these three studies (with the exception of *mdaB*), the aCGH results presented here support RD*holarctica*3 proposed by Broekhuijsen *et al.* which in turn were partially supported by Samrakandi *et al.*

4.3.5.4. RD*holarctica*4

Both Broekhuijsen *et al.* and Samrakandi *et al.* also proposed RD*holarctica*4: two ORFs were predicted by Broekhuijsen *et al.*: piperideine-6-carboxylate dehydrogenase (strongly homologous to aldehyde dehydrogenase) and a conserved hypothetical protein. Samrakandi *et al.* predicted only one ORF: piperideine-6-carboxylate dehydrogenase. Both of these proposed RD*holarctic*4 were mapped to FTT0552 (and FTT0553) of the genome sequence. In agreement with Samrakandi *et al.*, by arrayCGH RD*holarctica*4 was observed to consist of one ORF: FTT0552 which is annotated in the final version of the genome sequence as an aldehyde dehydrogenase. It was noted, however, that analysis of the oligonucleotide designed to bind FTT0553 using the basic local alignment search tool, nucleotide (BLASTn) (Altschul *et al.*, 1990), reveals 16 base pairs contiguous

homology elsewhere in the *F. tularensis* Schu S4 genome sequence, which is enough homology to allow cross-hybridization (Kane *et al.*, 2000).

4.3.5.5. RD*holarctica*5

The RD*holarctica*5 was also proposed by both Broekhuijsen *et al.* and Samrakandi *et al.*, although again there was a discrepancy between the two as to the exact ORFs included in the RD*holarctica*. Broekhuijsen *et al.* propose two ORFs: an adenine-specific DNA methyltransferase and a type I RM system. Samrakandi *et al.* predicted only the type I RM system. The proposed RD*holarctica*5 maps to the region of the *F. tularensis* Schu S4 chromosome containing the ORFs FTT0521 and FTT0522. By aCGH FTT0522 was found to be absent for three of the *F. tularensis* subsp. *holarctica* strains tested and marginal for another three. FTT0522 was present by arrayCGH for the three strains tested that originated from the former USSR, including the vaccine 'strain 15' FSC338. FTT0521 was not found to be absent in any of the strains tested although, BLASTn analysis of the relevant oligo revealed 16 base pairs contiguous homology elsewhere in the genome, possibly allowing cross-hybridization to occur. FTT0521 is annotated as a fragment in *F. tularensis* Schu S4, and was not included in the final genome sequence.

4.3.5.6. RD*holarctica*6

RD*holarctica*6 was proposed by Broekhuijsen *et al.* and by Samrakandi *et al.* Broekhuijsen *et al.* predict two ORFs: both hypothetical proteins which map to the ORFs *pdpD* and to the region between FTT1360 (*pdpD*) and

FTT1363 (ISFtu1, FTT1362 on the microarray) in the *F. tularensis* Schu S4 genome sequence. Samrakandi *et al.* predict three ORFs: *pdpD*, *vdcC*, and a conserved hypothetical protein. The final version of the *F. tularensis* genome does not include *vdcC*, but this maps to the area covered by the oligo designed to FTT1361. The conserved hypothetical protein maps to ORF FTT1362, ISFtu1. By aCGH all of the subsp. *holarctica* strains tested were deficient or marginal for *pdpD*. Six of the nine *F. tularensis* subsp. *holarctica* strains tested were absent or marginal in FTT1361 (putative *vdcC*). FTT1362 was absent only in the *F. novicida* strain tested, however as this oligonucleotide was designed to ISFtu1, found throughout the *Francisella* genome, it is impossible to determine whether this was as a result of a cross-hybridization.

4.3.5.7. $RD_{holarctica}7$

$RD_{holarctica}7$ was proposed by both Broekhuijsen *et al.* and Samrakandi *et al.*, although again there was a discrepancy between the annotations of the two. Broekhuijsen *et al.* propose two ORFs: an aspartokinase I, homoserine dehydrogenase and a homoserine kinase (possibly truncated). Samrakandi *et al.* predict only the aspartokinase I, homoserine dehydrogenase. The proposed $RD_{holarctica}7$ maps to the region of the *F. tularensis* chromosome containing the putative ORFs FTT0426 and FTT0427 (*thrA* and *thrB*, respectively). These two ORFs encode pseudogenes and, as such, were excluded from the annotated *F. tularensis* Schu S4 genome. By aCGH, *thrA* was absent or marginal from six of the nine *F. tularensis* subsp. *holarctica* strains tested.

4.3.5.8. RD*holarctica*8

RD*holarctica*8 proposed by Broekhuijsen *et al.* includes seven ORFs, which map from FTT1066 to FTT1072, flanked by ISFtu1 at FTT1065 and FTT1074. The same RD*holarctica* was proposed by Samrakandi *et al.* but with the addition of one ORF, FTT1073. This region was made up of hypothetical proteins and two pseudogenes (FTT1067 [*res*] and FTT1070) which have not been included in the *F. tularensis* Schu S4 genome sequence. Preliminary annotation of this region by both Broekhuijsen *et al.* and by Samrakandi *et al.* included the SNF2 family helicase *hepA*; a type III RM system; an antirestriction protein; and a *mobC*-like protein. Apart from the pseudogene *res*, which was preliminarily annotated as a type III RM system, none of the annotations proposed in the other studies have been carried through to the *F. tularensis* Schu S4 genome sequence. All of the European *F. tularensis* subsp. *holarctica* strains tested were absent or marginal for all of the ORFs that make up RD*holarctica*8. However, FSC012, the *F. tularensis* subsp. *holarctica* strain isolated from the USA did not conform to this pattern of deletion. This strain did have large deletion in this region of the chromosome, from FTT1071 to FTT1083.

4.3.5.9. RD*holarctica*9

Samrakandi *et al.* proposed RD*holarctica*9 as a hypothetical protein that mapped to FTT1308 in the *F. tularensis* Schu S4 genome sequence. The aCGH data presented here supported this finding in that FTT1308 was absent in all *F. tularensis* subsp. *holarctica* strains tested.

4.3.5.10. RD*holarctica*10

RD*holarctica*10 proposed by Samrakandi *et al.* partially overlaps with RD*holarctica*1: mapping to ORFs FTT1581-FTT1583. The results presented here support these findings in that FTT1582 and FTT1583 were absent or marginal in *F. tularensis* subsp. *holarctica* strains by aCGH.

4.3.5.11. RD*holarctica*11

RD*holarctica*11 was proposed by Samrakandi *et al.*, and by Svensson *et al.*, as FTT0531. The data presented here did not confirm this RD*holarctica*: this RD was identified only in FSC012, the only *F. tularensis* subsp. *holarctica* strain tested here that was isolated in the USA.

4.3.5.12. RD*holarctica*12

RD*holarctica*12 was proposed by Samrakandi *et al.* and maps to an insertion element, ISFtu2, in *F. tularensis* Schu S4. No probe for this was included on the *F. tularensis* microarray.

4.3.5.13. RD*holarctica*13

Samrakandi *et al.* proposed RD*holarctica*13 as an aminopeptidase and a membrane protein which mapped to FTT1789 and FTT1790 in the *F. tularensis* Schu S4 genome sequence. The aCGH data presented here did not support this finding for any of the strains tested, however the neighbouring ORF, FTT1791, was absent for all but one of the *F. tularensis* subsp. *holarctica* strains tested. The next downstream ORF, FTT1793, was absent for FSC012, the *F. tularensis* subsp. *holarctica* strain from the USA,

only. FTT1791 is annotated as a hypothetical protein and FTT1793 is *pepN*, an aminopeptidase.

4.3.5.14. RD*holarctica*15

RD*holarctica*15 was proposed by Samrakandi *et al.* as a reverse transcriptase that maps to FTT0921 in the *F. tularensis* Schu S4 genome sequence. FTT0921 is annotated as a hypothetical membrane protein but the aCGH results from this study support those of Samrakandi *et al.* in that all of the *F. tularensis* subsp. *holarctica* strains tested were absent or marginal for this ORF. In the final version of the *F. tularensis* strain Schu S4 genome sequence, there is no FTT0921, but FTT0920 is an ISFtu1.

4.3.5.15. Previously proposed RD*holarctica* not confirmed by these data

RD*holarctica*14, 16, 17, 18 (L3), 19 (L2), 20, 21, 22, and L1 also were proposed by Samrakandi *et al.*, or by Svensson *et al.*, or by both. The data from this study were unable to support these ORFs as RD*holarctica*.

4.3.5.16. New RD*holarctica* identified from this study

The following RD*holarctica* were identified from these data and have not been proposed in any previously published studies:

RD*holarctica*23, FTT0177, annotated as an acetyltransferase, was absent or marginal by aCGH for all subsp. *holarctica* strains tested, with the exception of FSC012, the *F. tularensis* subsp. *holarctica* strain isolated from the USA.

$RD_{holarctica}24$, FTT0446, a proton-dependent oligopeptide transport (POT) family protein. This ORF was absent or marginal in all of the *F. tularensis* subsp. *holarctica* strains tested.

$RD_{holarctica}25$ contains three ORFs, FTT0843-FTT0845. These were absent or marginal in all of the *F. tularensis* subsp. *holarctica* strains tested with the exception of FSC124, the strain isolated in the Ukraine, which was present for FTT0845. FTT0843 and FTT0844 are both annotated as pseudogenes in *F. tularensis* Schu S4, and FTT0845 is annotated as a hypothetical protein.

$RD_{holarctica}26$ was FTT1242, another hypothetical protein, and was absent or marginal in all of the *F. tularensis* subsp. *holarctica* strains tested.

$RD_{holarctica}27$ was FTT1426 which is annotated as a conserved hypothetical membrane protein in *F. tularensis* Schu S4, was absent or marginal in all of the *F. tularensis* subsp. *holarctica* strains tested (this ORF was excluded from analysis in FSC354 and FSC257).

$RD_{holarctica}28$ was FTT1670 (*napH*), a Na^+/H^+ antiporter fragment, and was absent or marginal from all of the *F. tularensis* subsp. *holarctica* strains tested.

4.3.6. RD_{LVS}

A comparison of RD identified by aCGH in FSC155 (*F. tularensis* LVS) and FSC338, the strain from which FSC155 is thought to have been derived (Tigertt, 1962) was made (fig. 4.14). There were 12 RDs additional to FSC155 (RD_{LVS}): FTT0118 (*prfC*); FTT0156; FTT0187 (*ftsA*); FTT0467 (*ostA1*); FTT0630 (*hfq*); FTT054 (*riml*); FTT1222 (*dedA2*); FTT1737 (*kdpC*); and oligo740. Unexpectedly there were also two RDs identified in FSC338, but not in FSC155, these were: FTT0221 (*acpA*); and FTT1277.

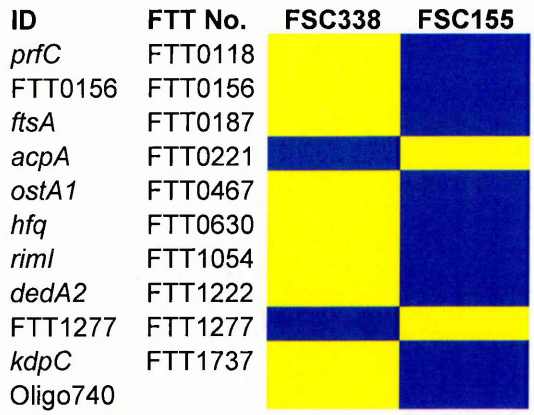


Fig. 4.14 A comparison of RD predicted in *F. tularensis* LVS by aCGH compared to RD predicted in the parent strain *F. tularensis* FSC338 by aCGH. RD identified in *F. tularensis* LVS but not in *F. tularensis* FSC338 were denoted RD_{LVS}.

Absent by aCGH

Present by aCGH

4.4. Discussion

4.4.1. aCGH studies on *Francisella*

A previous study by Svensson *et al.* (2005) sought to demonstrate the evolution of the subsp. of *F. tularensis* from the perspective of a series of unidirectional deletions (fig. 4.15). Briefly, it was proposed that *F. novicida* was the earliest to diverge from an ancestral strain through variation (but not deletion) of seven genes, followed later by the diversions of *F. tularensis* subsp. *holarctica* (through deletion of eight regions) and *F. tularensis* subsp. *mediaasiatica* (through deletion of one region). Broadly, the data presented here support this description in that the fewest number of RDs were identified in the *F. tularensis* subsp. *tularensis* strain tested, followed by *F. tularensis* subsp. *mediaasiatica*, with the greatest number of RDs observed in *F. tularensis* subsp. *holarctica*. These data also support the previous report that *F. tularensis* subsp. *mediaasiatica* is more similar to *F. tularensis* subsp. *tularensis* than *F. tularensis* subsp. *holarctica* (Sandström *et al.*, 1992). However, Svensson *et al.* did note that *F. novicida* was more divergent from *F. tularensis* subsp. *tularensis* than any of the *F. tularensis* subsp., as evidenced by the highest number of single nucleotide variations. The data for *F. novicida* presented in this study agree with that presented in the Svensson *et al.* study in that seven RDs were identified. However, as the microarray used here was not of suitable genome coverage to detect single nucleotide variations, the data from this study can neither be used to support nor refute the finding of Svensson *et al.* that *F. novicida* is overall most divergent from *F. tularensis* subsp. *tularensis*, although this is likely given the vast phenotypic differences between these two.

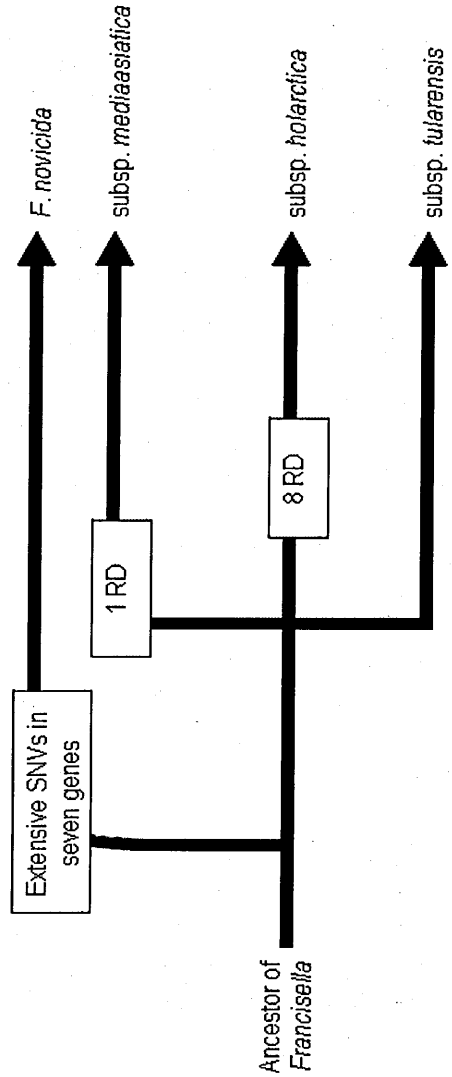
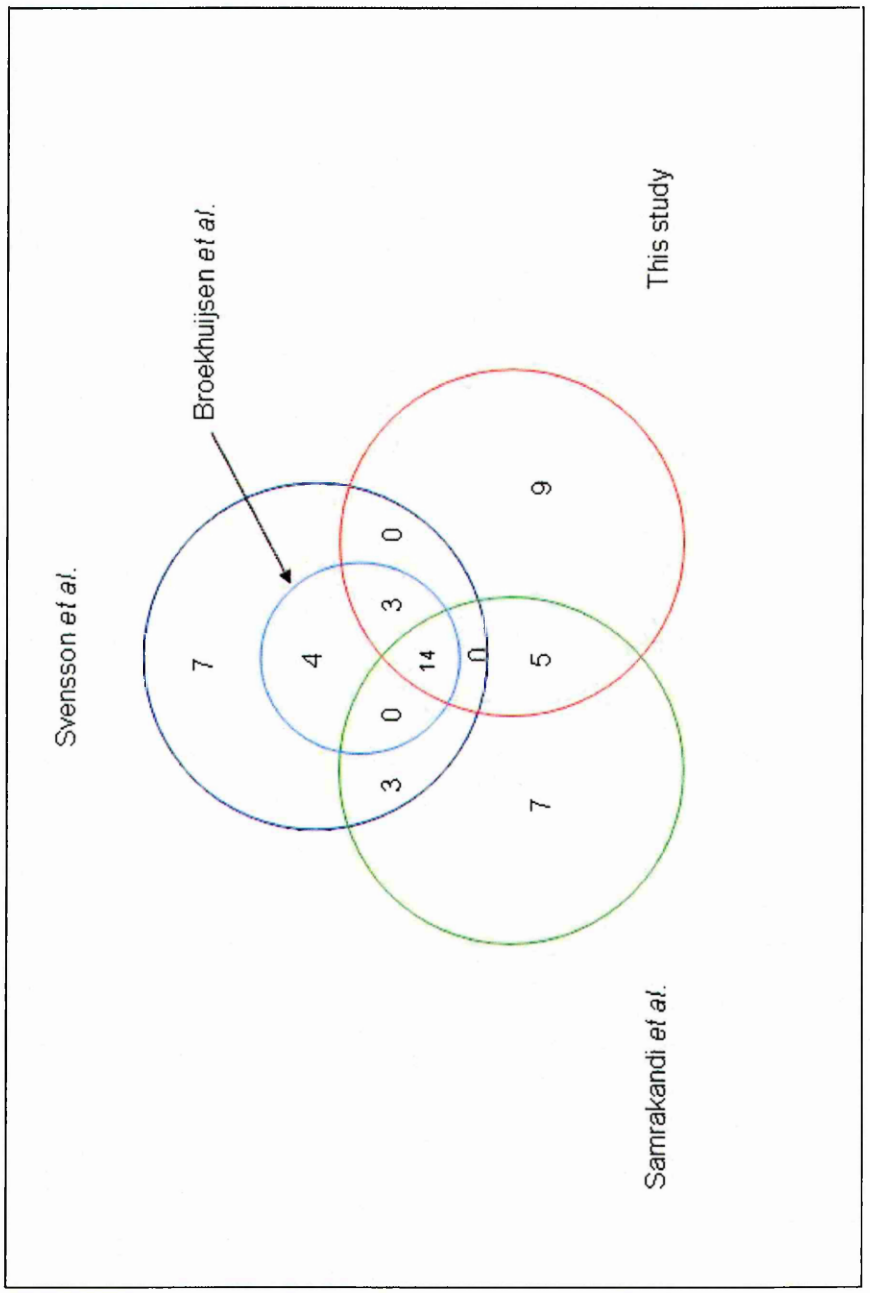


Fig. 4.15 An evolutionary scenario of *F. tularensis* adapted from Svensson *et al.* (2005). It was proposed that *F. novicida* was the earliest to diverge from the ancestral strain, followed by *F. tularensis* subsp. *holarctica* and *F. tularensis* subsp. *mediaasiatica*. The subsp. proposed as least divergent from the ancestral strain was *F. tularensis* subsp. *tularensis*.

The aCGH data from this study does not completely agree with previously published data, including the *F. tularensis* LVS genome sequence (Chain *et al.*, 2006), however, these data do not completely agree with each other either (fig. 4.16). For example, $RD_{holarctica}$ 1-4 were identified by Broekhuijsen *et al.*, and by Svensson *et al.*, as well as in this study, but $RD_{holarctica}1$ and 2 were not identified by Samrakandi *et al.*, (table 4.7). Additionally, the ORFs represented by $RD_{holarctica}1$-4 are present and intact in the published genome sequence of *F. tularensis* LVS. The fact that each of these studies has produced different yet overlapping data may indicate that some of the differences observed in each study represent genetic alterations that have occurred in the laboratory, as a result of culture or storage conditions. This hypothesis may be supported in that Svensson *et al.*, noted heterogeneity in two *F. tularensis* subsp. *holarctica* for the presence of $RD_{holarctica}$ 18 and 19, and suggested that this could be evidence that culture on laboratory media may facilitate fixation of deletion variants. That *F. tularensis* LVS is prone to genome alteration under laboratory conditions is also evidence of a certain degree of plasticity of the *Francisella* genome, not least because this vaccine strain was originally attenuated through continuous *in vitro* passage (Tigertt, 1962). It should not be forgotten that the evolution of this strain, and presumably all *Francisellae*, continues in the same way, a factor which could account for the slight variation in results observed between groups.

Fig. 4.16. A Venn diagram to show the number of ORFs coincident with $RD_{holarctica}$ reported in each study (Broekhuijsen *et al.*, 2003; Samrakandi *et al.*, 2004; Svensson *et al.*, 2005; this study). This diagram illustrates that a similar level of agreement exists between the data obtained from each of the four studies.


Svensson *et al.*
Broekhuijsen *et al.*
This study
Samrakandi *et al.*
7
4
3
0
14
0
3
0
9
5
7

4.4.2. $RD_{holarctica}$

As described in 4.3.5., a number of $RD_{holarctica}$, were identified using the *F. tularensis* Schu S4 microarray, and it is likely that one or more of these could account for the relative attenuation of *F. tularensis* subsp. *holarctica* compared to *F. tularensis* subsp. *tularensis*. The ORFs identified as $RD_{holarctica}$ fit into three categories with respect to the likelihood of the omission or deletion of these being responsible for the relative attenuation of *F. tularensis* subsp. *holarctica*: 1, the corresponding ORF in fully virulent *F. tularensis* Schu S4 is a pseudogene or the corresponding system is non-functional, making the absence of these ORFs unlikely to be responsible for attenuation; 2: the corresponding ORF is functional in *F. tularensis* Schu S4 but there is no previously reported role for the gene in virulence or pathogenesis. The absence of these ORFs *may* be responsible for the relative attenuation of *F. tularensis* subsp. *holarctica*; and 3: ORFs which correspond to genes for which there is a previously reported role in virulence or pathogenesis, either in *Francisella*, or in other bacteria. It is likely that the absence of one or more of these genes could be an attenuating factor of *F. tularensis* subsp. *holarctica*.

4.4.2.1 $RD_{holarctica}$ that are pseudogenes or part of redundant systems in *F. tularensis* Schu S4

The genes encoded in $RD_{holarctica}7$ (*thrA* and *thrB*), $RD_{holarctica}10$ (FTT1582 and FTT1583), and $RD_{holarctica}28$ (*napH*) are annotated as pseudogenes in *F. tularensis* Schu S4, making the absence of these genes unlikely to account for the relative attenuation of *F. tularensis* subsp. *holarctica* compared to *F. tularensis* subs. *tularensis*.

$RD_{holarctica}2$ has been identified as comprising *oppD* and *oppF* (Broekhuijsen *et al.*, date), or, in this study, just *oppF*. Both *oppD* and *oppF* are annotated as subunits of an oligopeptide transporter, an ATP-binding protein which is part of an ABC transporter system. The operon consisting of *oppABCDF* is a periplasmic binding protein-dependent transport system which handles peptides containing between two and five amino acids (Hiles *et al.*, 1987), the main function of the system being in recycling cell wall peptides (Hiles *et al.*, 1987; Perego *et al.*, 1991). However, the *F. tularensis* Schu S4 genome sequence reveals that *oppABC* are all pseudogenes. So, whilst *oppF* (and *oppD*) are annotated as functional genes in *F. tularensis* Schu S4, the lack of a functional *opp* system in this fully virulent *F. tularensis* subsp. *tularensis* strain probably means that the absence of *oppF* (or *oppD*) probably does not contribute to the reduced virulence of *F. tularensis* subsp. *holarctica*.

4.4.2.2 Annotated ORFs with no previously reported role in virulence

The gene encoded by *F. tularensis* Schu S4 in $RD_{holarctica}24$, FTT00446, is annotated as a proton-dependent oligopeptide transport (POT) family protein. This is one of at least eight POT family proteins (in addition to the non-functional *pot* system) that are encoded by *F. tularensis* Schu S4 (Atkins *et al.*, 2006), and it seems likely that one or more of these functions in *F. tularensis* subsp. *holarctica* for the uptake of oligopeptides, making the absence of FTT0446 unlikely to account for the relative attenuation of *F. tularensis* subsp. *holarctica*.

4.4.2.3 Annotated ORFs with a previously reported role in virulence

4.4.2.3.1 $RD_{holarctica}3$

This $RD_{holarctica}$ includes three ORFs encoding a conserved hypothetical protein, a modulator of drug activity, and a ThiJ/PfpI family protein. The ThiJ/PfpI protein superfamily also includes DJ-1, the human version of which shares a domain in common with *Escherichia coli* heat shock protein 31 (HSP31), this protein having both chaperone and proteolytic activities (Lee *et al.*, 2003). It has been suggested that, among other functions, DJ-1 may play a role in oxidative stress response (Tao and Tong, 2003). ThiJ genes are kinases that are involved in thiamine biosynthesis (Bandyopadhyay and Cookson, 2004; Oh *et al.*, 2004). In *Pyrococcus furiosus*, PfpI is an intracellular cysteine protease (Halio *et al.*, 1996) and has been identified in *Leishmania major*, which causes 'Old World' cutaneous leishmaniaisis, but not in other *Leishmania* species (Eschenlauer *et al.*, 2006). The potential role in infection of the modulator of drug activity, encoded by *mdaB*, is more obvious: MdaB is a soluble NADPH quinone reductase, a flavoprotein that catalyzes the two-electron reduction of quinone to quinols. It has been reported that MdaB protects cells from reactive oxygen species (ROS) that are formed during single electron redox reactions (Adams and Jia, 2005). This assertion may be supported by the finding that a Δ*mdaB* strain of *Helicobacter pylori* was significantly more sensitive to oxidative stress, and was significantly reduced in its ability to colonize mouse stomachs (Wang and Maier, 2004). It was reported by the same authors that homologues of MdaB exist in the genomes of many pathogenic bacteria (Wang and Maier, 2004). Therefore, although there is some published data

linking ThiJ/PfpI family proteins to virulence, without the complete identification of the encoded protein in *F. tularensis*, this evidence is too diffuse to infer that the absence of this gene in particular contributes to the relative attenuation of *F. tularensis* subsp. *holarctica*. However, the previously published data regarding the role of MdaB in virulence makes it more likely that the absence of this gene is an attenuating factor of *F. tularensis* subsp. *holarctica*. It could also be true that the absence of both of these genes, or all three genes including the hypothetical protein, contributes to the relative attenuation of *F. tularensis* subsp. *holarctica*, particularly in the light that there is some published evidence that the DJ1 protein from ThiJ/PfpI superfamily and MdaB may be similar in function as they both have been reported to play a role in the response to oxidative stress.

4.4.2.3.2 $RD_{holarctica}4$

This $RD_{holarctica}$ encodes an aldehyde dehydrogenase, a class of enzyme which has previously been linked to bacterial virulence: *aldA* from *V. cholerae* is included in a pathogenicity island that is associated with epidemic and pandemic strains, but is absent in non-toxic strains (Karaolis *et al.*, 1998). The expression of AldA was also found to be higher in *Klebsiella pneumoniae* strains that were isolated from patients with liver abscesses than from those without (Ma *et al.*, 2005). Is it therefore possible that a missing aldehyde dehydrogenase could be one of the attenuating factors that separate *F. tularensis* subsp. *holarctica* from *F. tularensis* subsp. *tularensis*?

4.4.2.3.3 RD*holarctica*6

This RD*holarctica* coincides with a previously described *Francisella* pathogenicity island (FPI) (Nano *et al.*, 2004). The *F. tularensis* Schu S4 genome shows that the two genes that comprise this RD encode a hypothetical protein and a pseudogene. A Δ*pdpD* strain of *F. novicida* has been shown to be defective for growth in macrophages and also defective for virulence in mice (Nano *et al.*, 2004). It is therefore possible that a lack of *pdpD* accounts in some part for the relative attenuation of *F. tularensis* subsp. *holarctica* compared to *F. tularensis* subs. *tularensis*. It is noted that this was one of two RD*holarctica* shared by *F. tularensis* subsp. *mediasiatica*, which is also of only moderate virulence compared to *F. tularensis* subsp. *tularensis* (Sandström *et al.*, 1992), further suggesting that *pdpD* may play an important role in the virulence of *F. tularensis* subsp. *tularensis*.

4.4.2.3.4 RD*holarctica*8

All of the genes included in RD*holarctica*8 that was observed for European strains are annotated in the *F. tularensis* Schu S4 genome sequence as hypothetical or pseudogenes (or both). RD*holarctica*8 identified from the only *F. tularensis* subsp. *holarctica* strain isolated in the USA tested was comprised of a different but overlapping set of ORFs (fig. 4.13), but again most of the genes included in this RD*holarctica* were annotated as hypothetical proteins in the *F. tularensis* Schu S4 genome sequence. The only two exceptions to this were two genes annotated as transcriptional regulators. One of these is designated *hipA*, hip standing for high persistence. This serine kinase has been linked to high persistence of a

minority of *E. coli* through the induction of bacteriostasis when faced with murein and DNA synthesis inhibitors (Moyed and Bertrand, 1983; Wolfson *et al.*, 1990; Korch and Hill, 2006). Bacteriostasis, usually induced as part of the SOS response, is a well characterized survival mechanism of bacteria that are under stress, so perhaps the absence of the *hipA* gene is one of the attenuating factors of this particular *F. tularensis* subsp. *holarctica* strain.

4.4.2.3.5 $RD_{holarctica}$ 23

In the *F. tularensis* Schu S4 genome sequence FTT0177 is annotated as an acetyl transferase, an enzyme which catalyses the transfer of acetyl groups from one compound to another. There are several examples of acetyl transferases acting as bacterial and yeast virulence factors: YopJ, a type III effector protein of *Y. pestis* which mediates the acetylation of threonine and serine residues that activate host MAPK kinase 6 (MAPPK6). Acetylation of MAPKK6 blocks phosphorylation and therefore inhibits activation of this immune signalling molecule (Mukherjee *et al.*, 2006). The *Candida albicans* GNA1 gene encodes glucosamine-6-phosphate acetyltransferase, and a DGNA1 strain is attenuated in a mouse model, with rapid clearance from host tissue (Mio *et al.*, 2000). In *L. monocytogenes*, activation of *virR*, which encodes a putative two-component response regulator that has been shown to be essential for virulence in mice, is through a mechanism involving variation in the level of phosphate acetylation (Mandin *et al.*, 2005). Again, it is noted that this was one of two $RD_{holarctica}$ shared by *F. tularensis* subsp. *mediasiatica*, again suggesting that the gene encoded by FTT0177 may play an important role in the virulence of *F. tularensis* subsp. *tularensis*.

4.4.3. RD_{LVS}

Attenuation of the *F. tularensis* vaccine strain, LVS, is widely believed to have been achieved by continuous laboratory subculture of the virulent *F. tularensis* subsp. *holarctica* strain that is currently denoted FSC338 (Tigertt, 1962). By comparing the RDs of *F. tularensis* LVS that are additional to the RDs of *F. tularensis* FSC338 (denoted RD_{LVS}) it may be possible to elucidate those ORFs whose deletion contribute to the attenuation of the vaccine strain. In the same way as the $RD_{holarctica}$ (4.4.2.), the ORFs identified as RD_{LVS} fall into three categories with respect to the attenuation of *F. tularensis* LVS: 1, the corresponding ORF in fully virulent *F. tularensis* Schu S4 is a pseudogene or the corresponding system is non-functional making the absence of these ORFs unlikely to be responsible for attenuation. Ideally the comparator for this category would be the pseudogene or system status of *F. tularensis* FSC338, but as the full genome sequence for this strain is not available at the time of writing, the status of the ORFs and systems in fully virulent *F. tularensis* Schu S4 was referenced instead; 2: the corresponding ORF is functional in *F. tularensis* Schu S4 but there is no previously reported role for the gene in virulence or pathogenesis. The absence of these ORFs *may* be responsible for the attenuation of *F. tularensis* LVS; and 3: genes for which there is a previously reported role in virulence or pathogenesis, either in *Francisella*, or in other bacteria. It is likely that the absence of one or more of these genes could be an attenuating factor of *F. tularensis* LVS.

4.4.3.1 RD_{LVS} that are pseudogenes or part of redundant systems in *F. tularensis* Schu S4

4.4.3.1.1. FTT1737 *kdpC*

The *kdp* potassium uptake system of *E. coli* is comprised of two separate operons: *kdpDE*, and *kdpABC*, in *F. tularensis* Schu S4 *kdpA* and *kdpE* are pseudogenes meaning that the *kdp* system is unlikely to be functional in this fully virulent strain. The aCGH probe for *kdpC* was a DHP in *F. tularensis* LVS but was present in FSC338, however as this system is probably non- functional in *F. tularensis* Schu S4, it is unlikely to contribute to the attenuation of the vaccine strain.

4.4.3.2 RD_{LVS} for which the ORF has no previously reported role in virulence

4.4.3.2.1. FTT0187 *ftsA*

The gene *ftsA* encodes a protein that is usually considered to be essential for cell division: it has been reported that *ftsA* deletion mutants of *Bacillus subtilis* are impaired for cell division (Beall and Lutkenhaus, 1992) and for sporulation (Kemp *et al.*, 2002). However, in *E. coli* it has also been reported that there are 12 proteins, including FtsA, that assemble at the cell midpoint to form the division septum, and that there may be some functional overlap between these 12, in particular between FtsA and another cell division protein denoted FtsK (Geissler and Margolin, 2005). Given that *F. tularensis* LVS cells are able to divide it may be that some similar compensatory mechanism for deletion of FtsA exists in this strain.

4.4.3.2.2. FTT0118 *prfC*, FTT1054 *riml*, FTT1222 *dedA2*

Termination of protein synthesis in bacteria requires the action of two codon-specific release factors (RF-1 and RF-2), the activities of which are stimulated by peptide chain release factor 3 (RF-3), encoded by *prfC* (Mikuni *et al.*, 1994). RF-3 is not thought to be essential for bacterial viability (Grentzmann *et al.*, 1994), with the disruption of *prfC* in *E. coli* leading to increased expression of the tryptophanase operon which in turn leads to the repression of tryptophan-induced inhibition of Rho-mediated transcription termination (Yanofsky *et al.*, 1996). The protein encoded by *riml* is an enzyme which, in *E. coli*, acetylates the N-terminal alanine of a the ribosomal protein S18 (Yoshikawa *et al.*, 1987). Acetylation of ribosomal proteins is required for the formation of the initiation complex during protein biosynthesis (Ramagopal and Subramanian, 1974). The DedA family of proteins are integral membrane proteins with an as yet un-elucidated function (Ledgham *et al.*, 2005). At the time of writing there is no evidence to suggest that deletion of any of these genes contributes to the attenuation of *F. tularensis* LVS.

4.4.3.3. RD_{LVS} for which the ORF has a previously reported role in virulence

4.4.3.3.1. FTT0156

Acid phosphatases have been implicated as virulence determinants of intracellular pathogens through suppression of the oxidative burst (Reilly *et al.*, 1996). *F. tularensis* encodes at least two acid phosphatases, *acpA* (FTT0221), and FTT0156, the latter being an RD_{LVS}. It has previously been

demonstrated that a Δ*acpA* mutant of *F. novicida* was attenuated in tissue culture assays: the mutant was delayed in phagosomal escape and was more susceptible to killing by a human monocyte cell line. The mutant also showed reduced longevity in mice in competition infection assays (versus wild type *F. novicida*), and mice which received the Δ*acpA* mutant survived for longer than those receiving wild type *F. novicida* (Mohapatra *et al.*, 2007). However, this report is in direct contrast to a previous study carried out using a *F. novicida* Δ*acpA* strain in which the mutant was reported to have exhibited wild type growth in a murine macrophage cell line and wild type replication in the murine model (Baron *et al.*, 1999). It is possible that the deletion of *acpA* in the Baron study was compensated for by the expression of the second acid phosphatase and that for some reason this effect was missing or was reduced in the Mohapatra study. Certainly an acid phosphatase has been shown by 2-D electrophoresis to be expressed in virulent strains of *F. tularensis* but not in *F. tularensis* LVS (Hernychová *et al.*, 2001). Together these findings demonstrate that acid phosphatase is probably important in *Francisella* infection, and it is possible to hypothesise that the absence of the acid phosphatase encoded by FTT0156 could be one of the attenuating lesions of the vaccine strain.

4.4.3.3.2. FTT0467 (*ostA*)

The protein encoded by *ostA* in Gram-negative bacteria is proposed as part of a targeting system for the transfer of outer membrane components from the site of biosynthesis (either in the cytoplasm or in the cytoplasmic leaflet of the inner membrane) across the inner membrane and periplasm to

the functional site (Braun and Silhavy 2002). Disruption of *ostA* in *H. pylori* results in altered membrane permeability, sensitivity to organic solvents and increased sensitivity to antibiotics (Chiu *et al.*, 2007). It is therefore possible that deletion of *ostA1* from *F. tularensis* LVS could also result in altered membrane permeability, perhaps making it more difficult for the vaccine strain to resist some of the components of the host defences that are encountered *in vivo*.

4.4.3.3.3. FTT0630 (*hfq*)

In *E. coli* hfq encodes a protein that can bind and stabilise small RNAs, while these in turn can regulate protein translation through binding with target mRNAs (Massé *et al.*, 2003). In *Yersinia enterocolitica* a gene homologous to *hfq* positively regulates expression of the virulence protein Yst (Nakao *et al.*, 1995), and in *S. enterica* serovar Typhimurium expression of the stationary phase transcription and virulence factor RpoS requires Hfq (Brown and Elliott, 1996). A *Brucella abortus* Δ*hfq* mutant showed increased sensitivity to hydrogen peroxide and decreased survival under acidic conditions during stationary phase growth. This mutant also failed to replicate in murine macrophages and was rapidly cleared from the spleens and livers of mice (Robertson and Roop, 1999). Similarly a *V. cholerae* Δ*hfq* mutant failed to colonize the intestine of a suckling mouse model (Ding *et al.*, 2004), and *hfq* has been shown to contribute to *L. monocytogenes* pathogenesis in mice (Christiansen *et al.*, 2004). It is therefore likely that *hfq* is required for the virulence that is displayed by *F. tularensis* subsp.

holarctica, and that the deletion of this gene from *F. tularensis* LVS may contribute to its attenuation.

4.5. Conclusion

This study has shown that the DstI *F. tularensis* Schu S4 microarray (chapter 3) is capable of discriminating between the species and subsp. of *Francisella* at the DNA level, with the data presented here being broadly similar to those previously published from similar studies. These data also support a description of the evolution of the *F. tularensis* subsp. proposed by Svensson *et al.* (2005). Furthermore this study has provided information about the genetic differences between fully and moderately virulent and avirulent strains of *Francisella*, which could prove to be useful in the search for medical countermeasures to tularemia, or in the provision of genetic targets for detection and identification at the subsp. level.

Chapter 5

In vitro stress responses of *F. novicida*

5.1. Introduction

Information about genes that are regulated by bacteria during infection can provide targets for vaccines, either through the generation of rationally attenuated strains or through the provision of potential antigens for subunit vaccines. Intracellular bacteria internalised in the phagosome (or autosome) of a host cell are often exposed to a variety of stress conditions, including extremes of acidity, oxidative stress, and nutrient deprivation. There are a great number of reports indicating that specific sets of genes are regulated, either up or down, by bacteria in response to the intracellular environment, and moreover that many of these *in* (or *ex*) *vivo*-regulated genes also show altered expression when the bacteria are exposed to stress conditions *in vitro*. For example, *Legionella pneumophila* was shown to induce at least 35 proteins during infection of the human monocyte cell line U937, 13 of which were also induced *in vitro* by one or more of the following stress conditions: heat shock, osmotic shock, and oxidative stress (Kwaik *et al.*, 1993). In the same study, 32 proteins were repressed by *L. pneumophila* during infection of U937 cells, of which 13 were also repressed by one or more of the stress conditions tested (Kwaik *et al.*, 1993). Similarly, the *L. monocytogenes* chaperone protein ClpC has been shown to be important for the bacteria's survival of osmotic and heat stress *in vitro*, and also for intracellular growth and for virulence (Rouquette *et al.*, 1998). Likewise, DnaK of *Brucella suis* is induced under conditions of heat shock and acidic pH, as well as upon infection of U937 cells (Köhler *et al.*, 1996). Of course, it is not true to say that all genes regulated by bacteria in response to the intracellular environment have also been observed to be regulated *in vitro*, for example a

study of *L. monocytogenes* gene expression upon infection of J774 murine macrophages showed the induction of 32 proteins, none of which were observed to induced by *L. monocytogenes* in response to heat shock or oxidative stress (Hanawa *et al.*, 1995), although selection of alternative *in vitro* conditions could yield different results.

It is because of the abundance of evidence suggesting overlapping bacterial responses to stress *in vitro* and to the intracellular environment that the *F. tularensis* microarray (chapter 3) was used to study the transcriptome of *F. novicida* in response to *in vitro* culture under conditions that were selected to resemble aspects of the intracellular environment faced by *Francisella* upon macrophage infection. *F. novicida* was used for these studies as it can be handled under ACDP containment level two conditions and it is often used as an analogue for highly virulent *F. tularensis*. The first condition investigated in this study was culture in an iron-depleted medium, in response to which a number of genes were regulated, both up and down. However, some of the genes regulated in response to this condition were indicative of a general stress response which may have been attributable to a difference in growth phase between the test and control samples at the time of sampling, as opposed to the specific response of growth in the absence of iron. In an attempt to address this issue and to distinguish genes regulated in response to the condition of interest from those regulated in response to growth phase variation, both test and control bacteria were cultured in the absence of methionine during the (parallel) sampling of the transcriptome under the remaining conditions. Culture without methionine arrested the growth of

both test and control cultures to the same level, meaning that changes observed in the transcriptome were attributable to the condition of interest as opposed to differences in growth phase. Each of the *in vitro* conditions studied: iron-depletion, oxidative stress, acidic pH, and elevated temperature, are presented here as separate sub-chapters.

Chapter 5a

The iron-starvation response of *F. novicida*

5a.1. Introduction

The importance of iron availability to intracellular pathogens is highlighted by the fact that individuals who suffer from the disease hereditary hemochromatosis, which is in part characterized by a lack of iron within macrophages, display increased resistance to intracellular pathogens, including *F. tularensis* (Moalem *et al.*, 1994). Iron is critical for bacterial growth as well as for host processes, it is an electron transport catalyst and a cofactor for a large number of enzymes that are required for many biological processes including aerobic metabolism (Escolar *et al.*, 1999; Zhang *et al.*, 2005). This makes iron-limitation inside phagosomes an effective mechanism that the host can use to inhibit bacterial survival and replication, meaning that only those bacterial species which have evolved ways to obtain and/or store iron will have a successful intracellular lifestyle.

The availability of iron in nature is limited because of rapid oxidation of ferrous iron (Fe^{2+}) to ferric iron (Fe^{3+}) under aerobic conditions and the low solubility of Fe^{3+} at physiological pH (Guerinot 1994, Braun *et al.*, 1998). In the host, any extracellular iron is bound to carrier proteins (siderophilins) such as transferrin which is the principle siderophilin in blood, or lactoferin, the principle siderophilin at mucosal surfaces (Neale, 1955; Ward *et al.*, 2003). The primary function of these proteins is to sequester iron and to transport this bound iron to tissues. Bacteria have evolved three broad mechanisms to obtain iron (Carroll *et al.*, 1996): 1) by causing tissue damage and utilising the resultant haem or haemin to acquire iron, for example *Shigella* species and *Streptococcus* species, 2) through direct binding of host

siderophilins to inducible proteins on the bacterial cell surface, for example *Neisseria* species., *Listeria* species., *Haemophilus* species., and *Staphylococcus* species., and 3) through the production of low molecular weight, non-protein iron binding molecules (siderophores) which have a higher affinity for iron than host siderophilins, for example *E. coli*. Siderophores are Fe^{3+} chelators that are secreted into the environment and are usually regulated by bacteria in response to iron levels (Tuanyok *et al.*, 2005). Ferri-siderophore complexes are then imported by transport proteins which are also regulated by levels of intracellular iron (Neilands, 1995). A siderophore has recently been identified and characterized in *F. tularensis* (Deng *et al.*, 2006; Sullivan *et al.*, 2006).

In the presence of H_2O_2 iron can participate in the formation of toxic oxygen radicals via the Fenton reaction, making regulation of iron uptake essential to prevent oxidative damage to the cell. The ferric uptake regulator (Fur) is a divalent metal ion-dependent DNA-binding protein that controls expression of genes involved in iron uptake in most Gram-negative bacteria (Bagg and Neilands, 1987). Fur proteins are generally transcription repressors by the action of ferrous-iron-dependent binding to specific sequences (iron or Fur boxes) encoded in promoter regions of iron-regulated genes (Escolar *et al.*, 1999). Positive regulation by Fur can also be achieved in either a direct or indirect manner. Fur is a pleiotropic regulator; it has been shown to control functions as diverse as iron acquisition systems, acid and oxidative stress responses, chemotaxis, swarming, bioluminescence, metabolic pathways, production of virulence factors, and even low-shear

modelled microgravity response (Vasil 2007; Ernst *et al.*, 2005; Ricci *et al.*, 2002; Wilson *et al.*, 2002).

5a.1.2. Aim

Determination of stress responses can provide clues as to the elusive survival and virulence mechanisms that *Francisella* employs during it's lifestyle as a highly adapted intracellular pathogen. As such, the objective of this study was to determine the pattern of gene expression as a defined response of *Francisella* to iron-starvation as a specific stress condition likely to resemble the macrophage environment during intracellular growth. The *F. tularensis* Schu S4 microarray was used to examine the *F. novicida* response to iron-starvation.

5a.2. Results

5a.2.1. Growth of *F. novicida* under iron-depleted conditions

In order to ascertain whether *F. novicida* was likely to be stressed by culture under iron-depleted conditions, and in order to select biologically relevant sampling times for transcriptomic studies, a comparison of growth rates under iron-depleted and iron-replete conditions was made. It was not until two hours after stress induction that a difference between the growth rates was observed in iron-replete and iron-depleted cultures of *F. novicida* (fig. 5a.1). Presumably this is because the bacteria retain some intracellular reserves of iron which can be used to sustain vital cellular functions and it is once these stocks become very low or exhausted that an affect on growth rate is observed, which is an indication of the importance of iron to the cell.

5a.2.2. The transcriptomic response of *F. novicida* to iron-depleted conditions

The transcriptomic response of *F. novicida* to culture under iron-depleted conditions was investigated using a DNA microarray. Based on the comparative growth rates (fig. 5a.1), the following time points post stress-induction were selected for transcriptomic analysis: at 1 h, when the growth rates of *both* test and control cultures were similar, at 2 h, when the growth rates the test and control culture had just diverged, and at 3.5 h, when the growth rates of the test and control cultures were fully divergent and the control culture was in log phase. RNA was isolated for each condition on three separate occasions (biological replicates) and hybridised on two separate occasions (technical replicates).

A list of genes that were up- and down-regulated in response to iron starvation can be found in tables 5a.1 and 5a.2. A number of genes were regulated in response to iron starvation after just one hour, before the growth rate was visibly affected by depletion of iron, which may be a further indication of the importance of iron to *F. novicida*.

5a.2.2.1. Some regulated oligonucleotides were not complementary to *F. novicida* CDS

The microarray oligonucleotides used to carry out *F. novicida in vitro* stress response studies were originally designed to be complementary to the *F. tularensis* Schu S4 genome sequence. With this in mind, the sequence of each oligonucleotide that was identified as regulated by *F. novicida* was compared to the *F. novicida* genome sequence that was subsequently available (Brittnacher *et al.*, 2006) using the BLASTn program (Altschul *et al.*, 1990). In most cases 100% complementarity was achieved between the oligonucleotide and the *F. novicida* genome sequence and within the *F. novicida* CDS corresponding to the *F. tularensis* Schu S4 CDS to which the oligonucleotide was designed. If less than 100% complementarity between the oligonucleotide and the *F. novicida* genome was achieved, then the most likely *F. novicida* binding site was selected according to the criteria of ≥ 75% overall complementarity and ≥ 15 base pairs contiguous complementarity (Kane *et al.*, 2000). In these cases the sequence of the *F. novicida* CDS selected was then used to find the corresponding CDS in *F. tularensis* Schu S4 using tBLASTx (Altschul *et al.*, 1990) to ensure continuity when interpreting the results in the context of either genome. In the case of

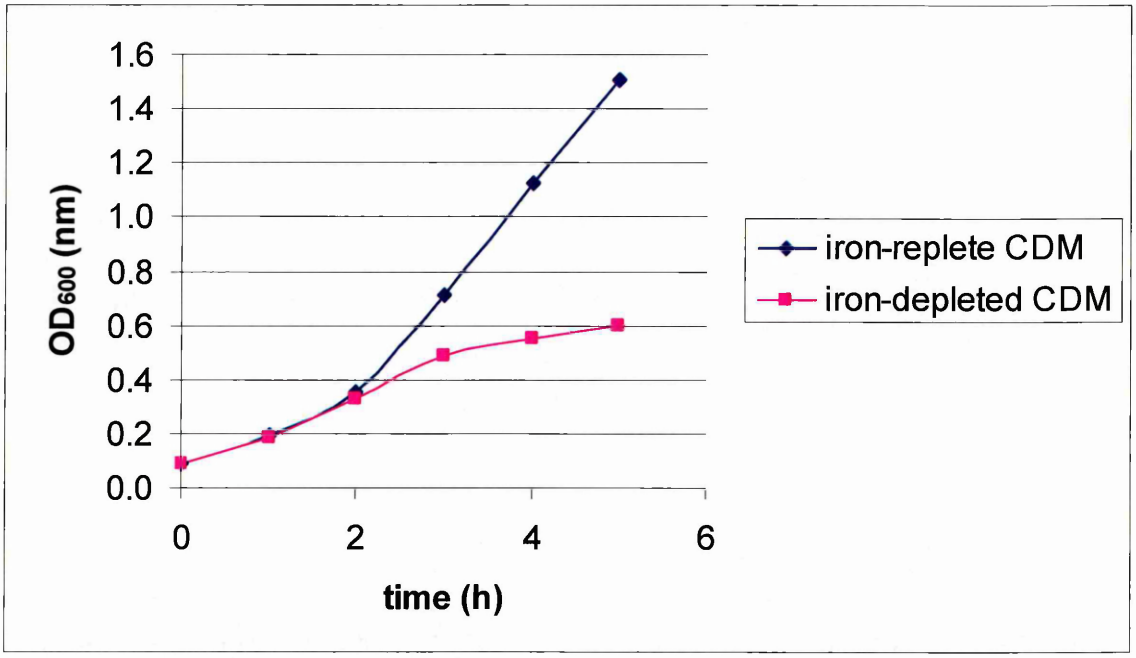


Fig. 5a.1 Comparison of the replication rate of *F. novicida* cultured in CDM under iron-depleted vs. iron-replete conditions. The growth rate was equivalent for both test and control cultures until two hours post stress induction, when the growth rate of *F. novicida* cultured under iron-starved conditions was reduced.

Table 5a.1 Genes up-regulated by *F. novicida* in response to culture under iron-deficient conditions compared to iron-replete conditions. [a]Locus tag assigned in the *F. novicida* U112 genome sequence (Brittnacher *et al.*, 2006). [b]Locus tag assigned in the *F. tularensis* Schu S4 genome sequence (Larsson *et al.*, 2005). [c]Pseudogene in *F. tularensis* Schu S4.

ID	FTN No.[a]	*FTT no.*[b]	Time	Fold change	P_{FDR}	Annotation	GenProtEC classification
feoB	0066	*FTT0249*	60	3.54	0.04	Ferrous iron transport protein B	[4.9.A.8] [5.5.7] [6.1] [7.3]
pilV	0413	*FTT0888*	60	1.51	0.04	Type IV pili, pilus assembly protein	[1.6.13] [6.5]
FTN_1125	1125	*FTT1144*[c]	60	1.46	0.04	Short-chain dehydrogenase	[12]
lysA	1530	*FTT0027*	60	4.91	0.05	Diaminopimelate decarboxylase	[1.5.1.7] [11]
FTN_1685	1685	*FTT0026*	60	7.55	0.04	Drug H^+ antiporter-1 (DHA1) family protein	[4.2.A]
dnaN	0002	*FTT0002*	120	1.67	0.04	DNA polymerase III, beta subunit	[2.1.1] [7.1]
pyrF	0035	*FTT1648*	120	1.47	0.05	Orotidine-5'-phosphate decarboxylase	[1.5.2.2] [11]
ostA2	0713	*FTT0740*	120	1.55	0.04	Organic solvent tolerance protein OstA	[5.5.6] [6.1] [7.3]
FTN_0753	0753	*FTT0580*	120	1.50	0.05	Hypothetical protein	[13]
FTN_0755	0755	*FTT0582*	120	1.67	0.03	4Fe-4S Ferredoxin	[12]
sufB	0851	*FTT0971*	120	2.77	0.04	SufS activator complex, SufB subunit	[2.3.4] [7.1]
ggt	1159	*FTT1181*	120	1.61	0.04	Gamma-glutamyltranspeptidase	[1.5.3.10]
FTN_1169	1169	*FTT1191*	120	1.47	0.04	Peptidase M20 family	[12]
bfr	1410	*FTT1441*	120	1.76	0.05	Bacterioferritin	[5.5.7] [7.1]
lysA	1530	*FTT0027*	120	8.24	0.04	Diaminopimelate decarboxylase	[1.5.1.7] [11]
FTN_0979	0979	*FTT0536*	210	2.45	0.05	Major facilitator superfamily (MFS) transport protein	[4.2.A.1]
pyrD	0036	*FTT1647*	210	1.65	0.05	Dihydroorotate dehydrogenase	[1.5.2.2] [7.1] [11]
FTN_0362	0362	*FTT0847*	210	1.47	0.05	Deoxyribodipyrimidine photolyase-related protein	[12]
FTN_0721	0721	*FTT0749*	210	1.46	0.05	Hypothetical protein	[13]
coaD	0754	*FTT0581*	210	1.45	0.04	Phosphopantetheine adenylyltransferase	[1.5.3.5] [6.3]
sufB	0851	*FTT0971*	210	4.97	0.04	SufS activator complex, SufB subunit	[2.3.4] [7.1]
rubA	1084	*FTT0595*	210	1.51	0.04	Rubredoxin	[12]
FTN_1612	1612	*FTT0103*	210	1.48	0.05	Hypothetical protein	[13]
frgA	1682	*FTT0029*	210	9.60	0.04	Siderophore biosynthesis protein	[11]
FTN_1683	1683	*FTT0028*	210	3.82	0.05	Drug H^+ antiporter-1 (DHA1) family protein	[4.2.A] [11]
FTN_1685	1685	*FTT0026*	210	10.51	0.04	Drug H^+ antiporter-1 (DHA1) family protein	[4.2.A]
trpB	1739	*FTT1773*	210	1.64	0.05	Tryptophan synthase beta chain	[1.5.1.15] [7.1]

Table 5a.2 Genes down-regulated by *F. novicida* in response to culture under iron-deficient conditions compared to iron-replete conditions. [a]Locus tag assigned in the *F. novicida* U112 genome sequence (Brittnacher *et al.*, 2006). [b]Locus tag assigned in the *F. tularensis* Schu S4 genome sequence (Larsson *et al.*, 2005). [c]Pseudogene in *F. tularensis* Schu S4. [d]Approximately 80% from the 3'-end of the nucleotide sequence of FTN_0009 is homologous to the intergenic region between *sdaC2* and an isftu1 transposase in *F. tularensis* Schu S4.

ID c	FTN No.[a]	*FTT no.*[b]	Time	Fold change	P_{FDR}	Annotation	GenProtEC classification
FTN_1256	1256	*FTT1238*	60	1.54	0.05	Membrane protein of unknown function	[11] [12]
FTN_0009	0009	*None*[d]	120	1.49	0.05	Hypothetical protein	[13]
orn	0068	*FTT0228*	120	1.48	0.04	Oligoribonuclease (3'-5' exoribonuclease)	[2.2.4]
FTN_0326	0326	*FTT1610*[c]	120	1.46	0.05	Hypothetical protein	[13]
FTN_0747	0747	*FTT0574*[c]	120	1.47	0.04	Amino acid-polyamine-organocation (APC) superfamily protein	[4.2.A.3]
FTN-_0894	0894	*FTT1016*	120	1.47	0.05	GDSL-like lipolytic enzyme	[12]
nadE	1278	*FTT1259*	120	1.67	0.05	NAD synthase	[1.5.3.7] [1.7]
wbtE	1426	*FTT1460*	120	1.50	0.05	UDP-glucose/GDP-mannose dehydrogenase	[1.6.3.1] [6.3] [11]
rnc	1463	*FTT1555*	120	1.45	0.03	Ribonuclease III	[1.2.1] [2.2.4]
FTN_1609	1609	*FTT0106*	120	1.48	0.04	Membrane fusion protein	[12]
hiss	1658	*FTT0052*	120	1.51	0.05	Histidyl-tRNA synthetase	[2.3.1] [7.1] [11]
FTN_1665	1665	*FTT0046*[c]	120	1.48	0.04	Magnesium chelatase	[1.7.19]
clpB	1743	*FTT1769*	120	1.55	0.04	Chaperone ClpB	[1.2.3] [2.3.4] [7.1] [11]
FTN_0326	0326	*FTT1610*[c]	210	1.67	0.05	Hypothetical protein	[13]
glgP	0517	*FTT0417*	210	1.59	0.04	Glycogen phosphorylase	[1.2.4.glycogen] [1.6.9] [7.1]
ftsI	0607	*FTT0697*	210	1.47	0.05	Cell division protein, peptidoglycan synthetase (PBP)	[1.6.7] [5.1] [5.6.4] [6.1] [6.2] [7.3]
yajC	1096	*FTT1116*	210	1.58	0.05	Preprotein translocase, subunit D, membrane protein	[4.3.A.5] [6.1] [7.3]
pdpD	1325	*FTT1360*	210	1.45	0.04	Hypothetical protein	[11] [13]
wbtH	1421	*FTT1456*	210	1.66	0.04	Glutamine amidotransferase/ asparagine synthase	[1.6.3.1] [6.3] [11]
FTN_1455	1455	*FTT1546*[c]	210	1.50	0.05	Hypothetical protein	[13]
rnr	1461	*FTT1553*	210	1.51	0.04	Ribonuclease R	[1.2.1] [2.2.4]
rnc	1463	*FTT1555*	210	1.55	0.04	Ribonuclease III	[1.2.1] [2.2.4]
FTN_1511	1511	*FTT1501*[c]	210	1.47	0.05	Hypothetical protein	[13]
tolC	1703	*FTT1724*	210	1.46	0.03	Outer membrane efflux protein, TolC precursor	[4.1.B] [5.1] [6.1] [7.4]
kdpD	1715	*FTT1736*	210	1.46	0.04	Two component regulator, sensor histidine kinase	[3.1.2] [6.1] [7.3] [2.3.3] [3.1.3.2] [11]
FTN_1736	1736	*FTT1776*	210	1.49	0.04	Hypothetical protein	[13]

regions of acceptable complementarity outside of CDS in *F. novicida* for two of the regulated oligonucleotides (table 5a.3).

5a.2.3. Regulated CDS were grouped according to function

Each of the CDS regulated by *F. novicida* in response to iron-depletion was assigned to a functional category as previously defined in the GenProtEC database (Serres *et al.*, 2004) (Appedix 1), and represented graphically (fig. 5a.2). The GenProtEC classification assigns gene products to groups according to their broad function, and into sub-groups which describe increasingly specific functions at each level. It should be noted that some CDS are assigned to more than one GenProtEC category.

5a.2.3.1. Functional groups of up-regulated CDS

The functional group containing the highest number of up-regulated CDS was Metabolism (fig. 5a.3). Within this functional group, all of the up-regulated CDS were assigned to biosynthesis, either macromolecular biosynthesis (fimbria, pili), or building block biosynthesis (amino acids, nucleotides, and cofactor or small molecule carriers). The second largest functional group of up-regulated CDS was Pathogenic-Related Genes, followed by Transport and Conserved Hypothetical groups. Of the eight up-regulated CDS for which a cellular location of the final gene product was assigned, six were located in the cytoplasm, and two were located in the inner membrane (fig. 5a.4). Four up-regulated CDS were assigned to a cell structure sub-category: two were assigned to the Membrane, one to Surface Antigens, and one to Pilus (fig. 5a.5). Of the three up-regulated CDS

Table 5a.3 Oligonucleotides not complementary to *F. novicida* CDS. [a]Locus tag assigned in the *F. tularensis* Schu S4 genome sequence (Larsson *et al.*, 2005). [b]Oligonucleotides were designed according to a preliminary annotation of the *F. tularensis* Schu S4 genome sequence. The putative CDS against which this oligonucleotide was designed was not included in the final version of the *F. tularensis* Schu S4 genome sequence.

Oligonucleotide	**FTT no.[a]**	**Time point**	**Fold change**	**P_{FDR}**	***F. novicida* locus**
Oligo754	None[b]	120 min.	1.47-fold up	0.03	Intergenic region between FTN_0591 and FTN_05892
Oligo1951	0277	120 min.	4.53-fold down	0.05	Intergenic region between FTN_0191 and FTN_0192

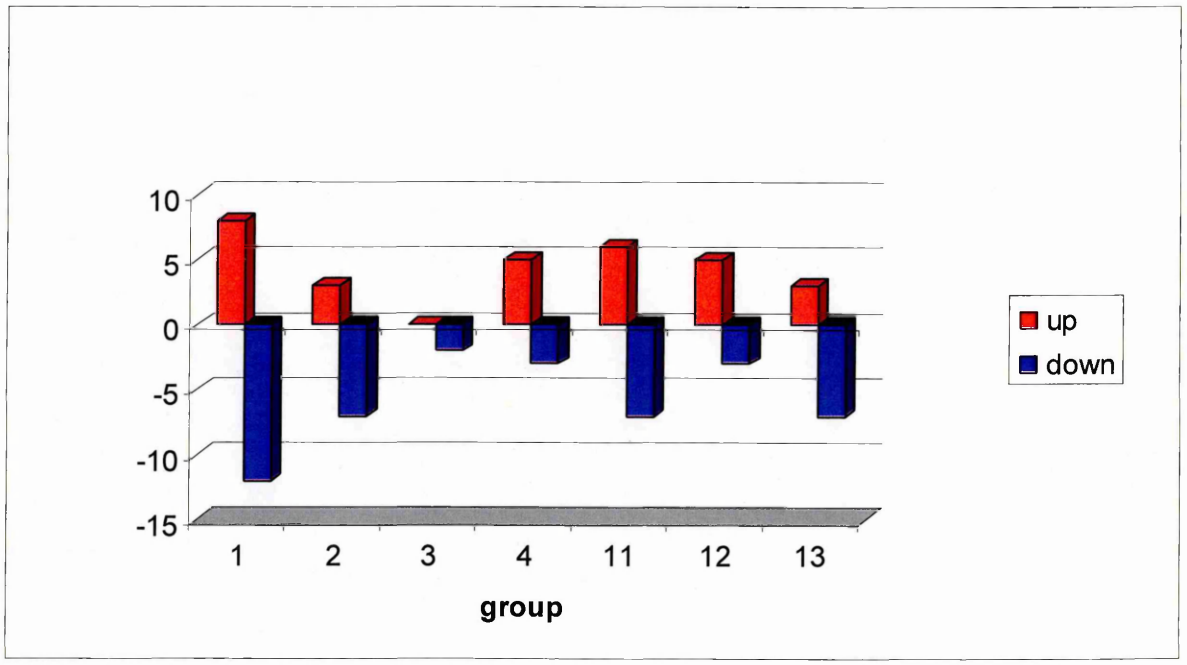


Fig. 5a.2 GenProtEC categories of CDS regulated by *F. novicida* in response to culture under iron-deficient compared to iron-replete conditions. The x-axis represents GenProtEC categories of CDS regulated in response to iron-depletion were Metabolism (category 1), Information transfer (category 2), Regulation (category 3), Transport (category 4), Pathogenic related genes (category 11), Conserved hypothetical genes (category 12), and Hypothetical genes (category 13), and the y-axis represents the number of genes up-regulated (+) or down –regulated (-).

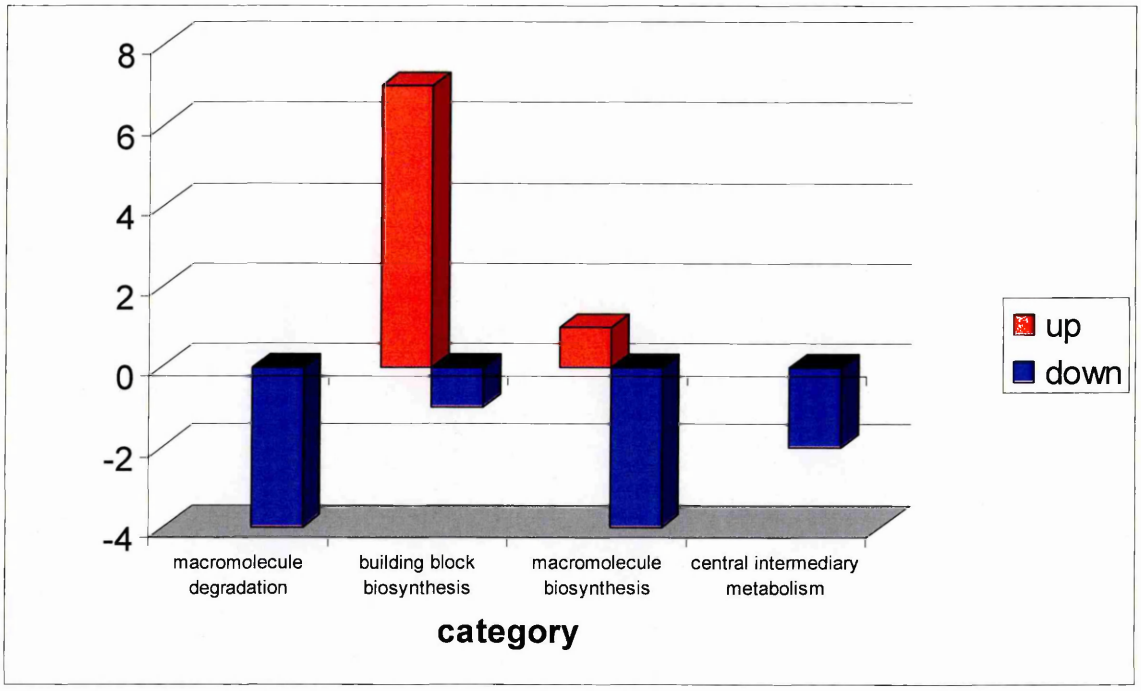


Fig. 5a.3 GenProtEC Metabolism sub-categories of CDS regulated by *F. novicida* in response to culture under iron-deficient compared to iron-replete conditions (x-axis). The y-axis represents the number of genes up-regulated (+) or down –regulated (-).

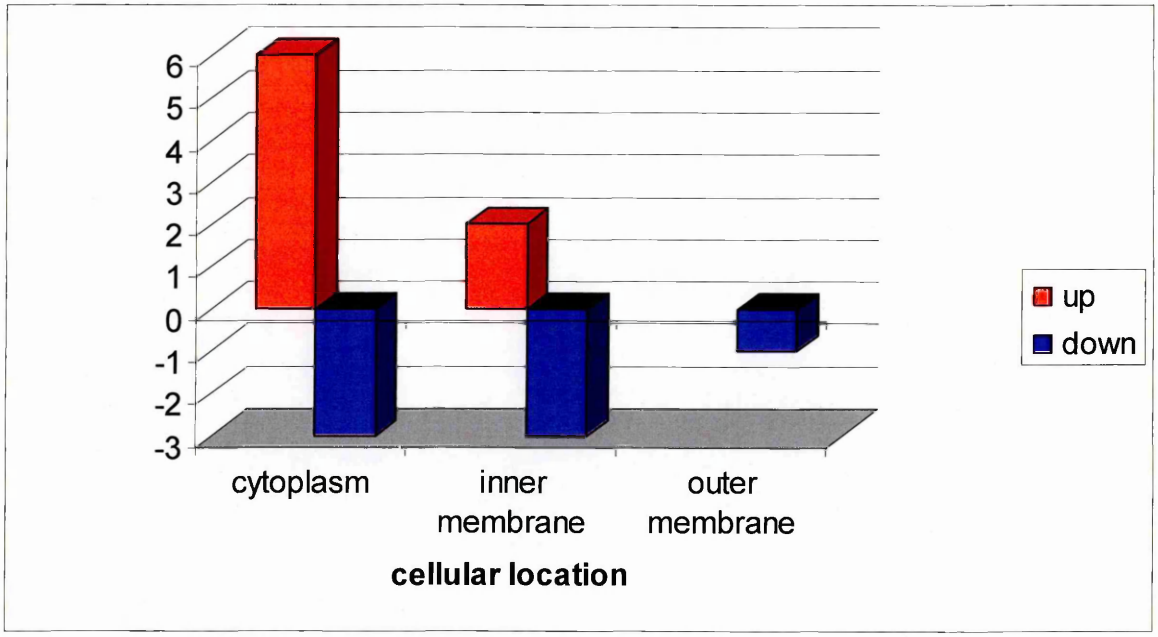


Fig. 5a.4 GenProtEC Cellular locations of CDS regulated by *F. novicida* in response to culture under iron-deficient compared to iron-replete conditions (x-axis). The y-axis represents the number of genes up-regulated (+) or down –regulated (-).

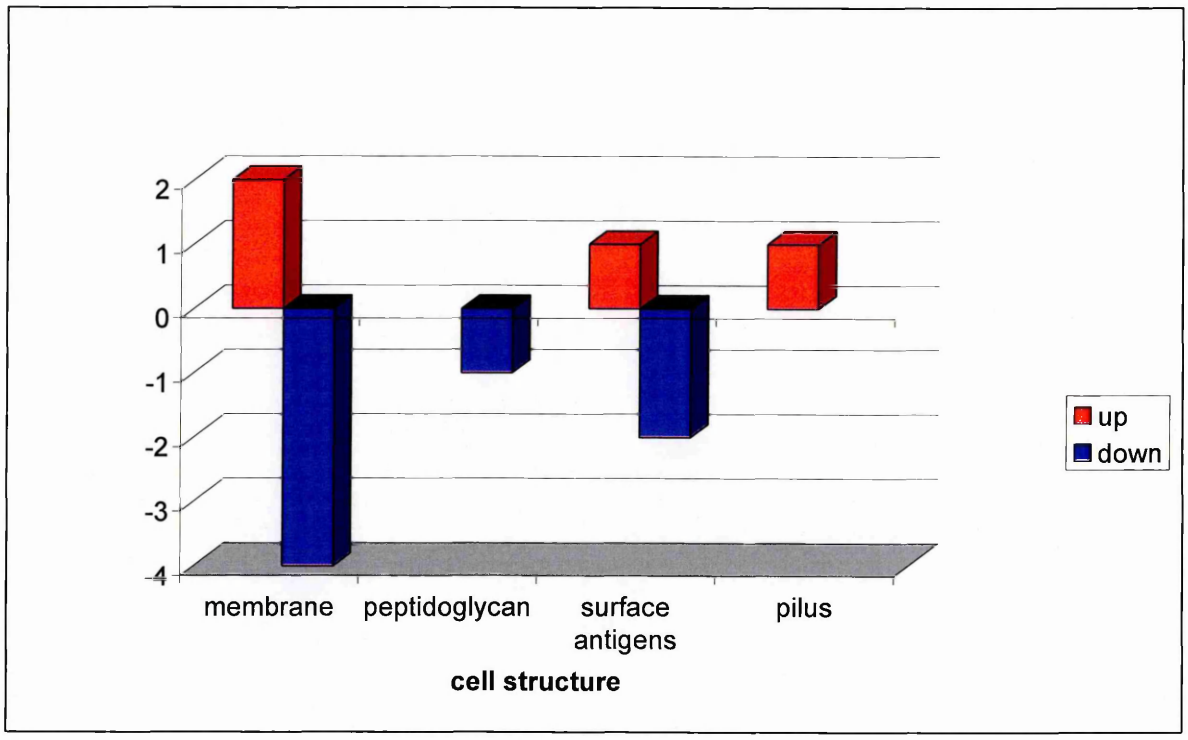


Fig. 5a.5 GenProtEC Cell structure-associated CDS regulated by *F. novicida* in response to culture under iron-deficient compared to iron-replete conditions (x-axis). The y-axis represents the number of genes up-regulated (+) or down –regulated (-).

assigned to the Cell Processes category, all were assigned to the sub-category for Adaptation to Stress (Iron Acquisition or 'Other') (fig. 5a.6).

5a.2.3.2. Functional groups of down-regulated CDS

The functional group containing the largest number of down-regulated CDS was also Metabolism, although these CDS covered a wider range of sub-categories than the up-regulated group, including: macromolecule degradation (RNA, polysaccharides and proteins/peptides/glycopeptides), building block biosynthesis (nicotinamide adenine dinucleotide [NAD]), macromolecule biosynthesis (O-antigen, peptidoglycan, and cytoplasmic polysaccharide), and central intermediary metabolism, including incorporation of metal ions (fig. 5a.2). The functional groups with the second largest groups of down-regulated CDS were Information Transfer (including both RNA- and protein-related groups), Pathogenic-related genes, and Hypothetical genes. Of the seven down-regulated CDS for which a cellular location of the final gene product was assigned, three were located in the cytoplasm, three were located in the inner membrane, and one was located in the outer membrane (fig. 5a.4). Seven down-regulated CDS were also assigned to a Cell Structure sub-category: four were assigned to the Membrane, two to Surface Antigens, and one to Peptidoglycan (fig. 5a.5). Of the three down-regulated CDS assigned to the Cell Processes category, two were assigned to Cell Division, and one was assigned to Protection (Drug Resistance/sensitivity) (fig. 5a.6).

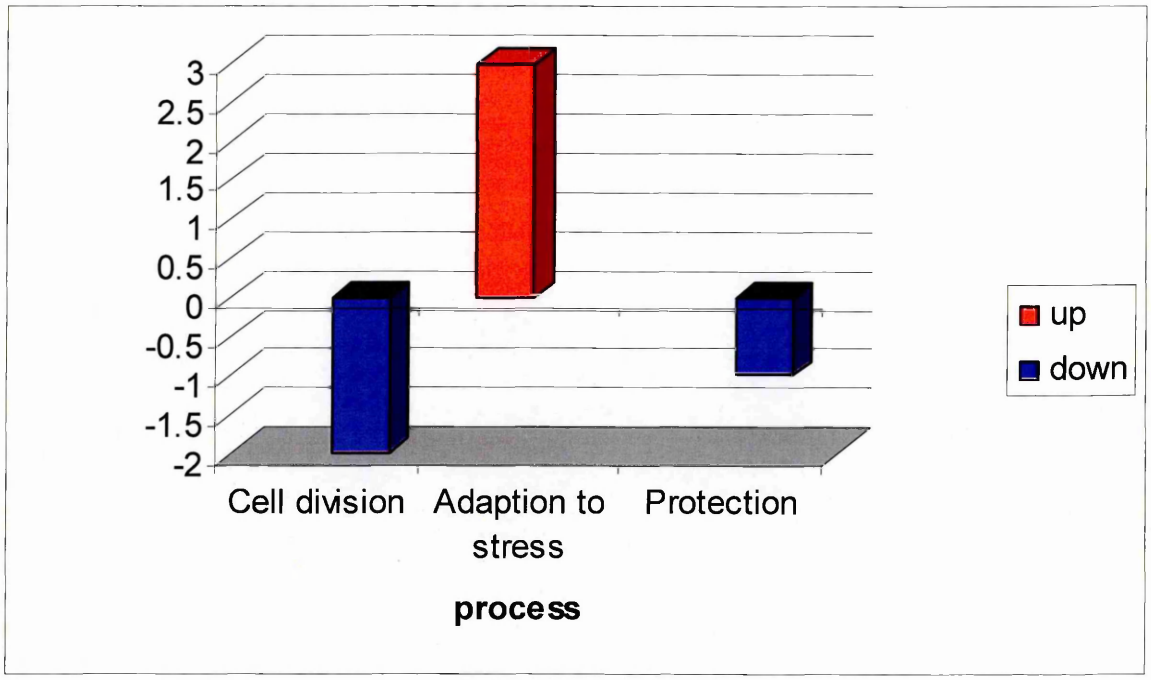


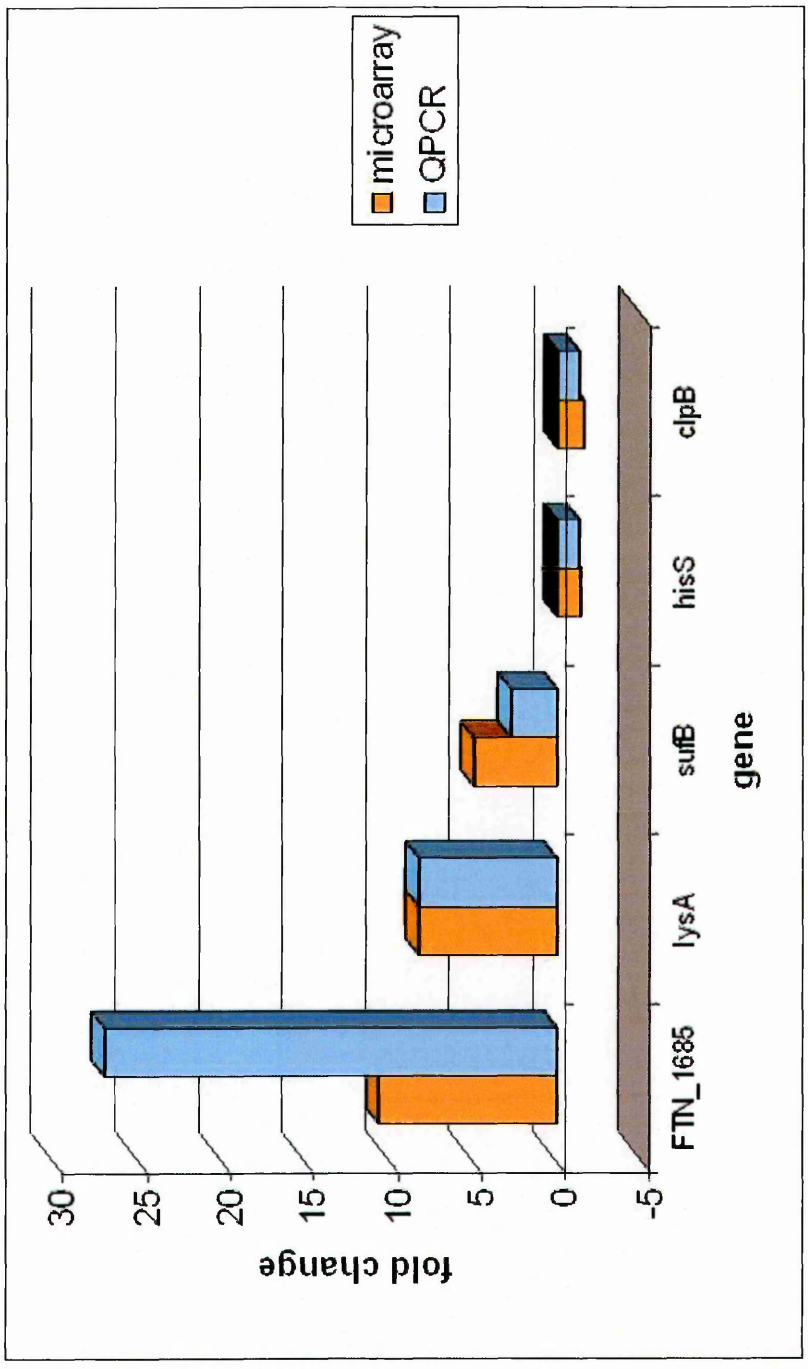
Fig. 5a.6 GenProtEC Cell process-associated CDS regulated by *F. novicida* in response to culture under iron-deficient compared to iron-replete conditions (x-axis). The y-axis represents the number of genes up-regulated (+) or down –regulated (-).

5a.2.4. QPCR

TaqMan QPCR was carried out on five CDS that were indicated by microarray as up-regulated by *F. novicida* in response to iron-starvation: FTT0026, FTT0971, and *lysA* were up-regulated (table 5a.1), and *hisS* and *clpB* were down-regulated (table 5a.2). Assays were performed and the results calculated as described in section 2.7. These microarray results were confirmed by QPCR with a Pearson correlation of 0.730 (fig. 5a.7). The overall Pearson correlation between microarray data and TaqMan QPCR data for 18 genes assayed, selected because they were indicated by microarray as regulated by *F. novicida* in response to one of the conditions tested was 0.533.

Fig. 5a.7 QPCR confirmation of microarray data. The QPCR data supported the microarray data in that all genes were regulated in the same direction, although not always to the same magnitude:

Gene	Microarray	QPCR
FTN_1685	10.51	27.03
lysA	8.24	8.2
sufB	4.97	2.77
hisS	-1.51	-1.33
clpB	-1.55	-1.36
Correlation	**0.730070558**	


microarray
QPCR
fold change
30
25
20
15
10
5
0
-5
FTN_1685
lysA
sufB
hisS
clpB
gene

5a.3 Discussion

Apart from hypothetical and conserved hypothetical proteins, all of the CDS up-regulated by *F. novicida* in response to iron-depletion fell into one of four GenProtEC classification groups: Metabolism, Transport, Information transfer, or Pathogenesis. Metabolic systems that were putatively up-regulated included peptidoglycan synthesis, the up-regulation of which may be linked to the SOS response, and cofactor or small binding molecule synthesis, which was linked to a putative iron acquisition system. Transport systems that were up-regulated included genes for the acquisition and transport of iron, and some Information transfer categorised genes that were up-regulated were linked to iron-sulphur cluster repair. Some groups of regulated genes with clearly corresponding functionality were actually assigned to separate (and sometimes multiple) GenProtEC groups, serving as a warning against relying solely upon sometimes somewhat arbitrary systems to elucidate biological meaning from data. One example of this was the up-regulation of genes that have previously been shown to be involved in the biosynthesis of a *Francisella* siderophore and which were assigned to three GenProtEC functional groups: Metabolism, Transport and Pathogenic-related genes.

The CDS down-regulated by *F. novicida* in response to iron-depletion fell into the same four GenProtEC categories as those that were up-regulated: Metabolism, Transport, Information transfer, or Pathogenic-related genes, with one additional category, Regulation, also represented. Metabolic systems that were down-regulated included LPS O-antigen biosynthesis and

incorporation of metal ions, which is not unexpected under conditions of iron ion starvation. Transport systems that were down-regulated included an efflux pump, which is hypothesised here to transport siderophores out of the cell. Several ribonuclease genes were also down-regulated (categorised in the GenProtEC groups Metabolism and Information transfer), as well as several genes with a previously demonstrated role in *Francisella* virulence, which may indicate that *F. novicida* does not encounter conditions of iron-depletion *in vivo*.

It was noted that, with the exception of the siderophore biosynthesis genes, no other complete operons were regulated by *F. novicida* in response to iron-starvation. This is an indication that the microarray data was noisy, resulting in a reduced number of genes with a p-value below 0.05 which was required to be counted as regulated. One suggestion as to the cause of this noise is the use of an air-incubator for culture as it has been suggested that the temperature is actually highly variable across the inside of the incubator. The temperature is not so variable as to cause detectable differences in growth rate, but is thought to be noticeable at the level of gene regulation. A better alternative for microarray work would be a water incubator, or ideally, a chemostat, as this would keep every aspect of the culture conditions, including the growth phase, constant. This trend of noisy data resulting in reduced numbers of regulated genes was noticed in all of the *in vitro* conditions investigated here.

5a.3.1. The reduced growth-rate may be linked to the SOS response.

A decreased growth rate was observed in *F. novicida* after two and a half hours culture under iron-depleted conditions compared to iron-replete conditions (section 5a.1), which may in part be attributed to the SOS response. The bacterial SOS response aids propagation in part by ceasing cell division at times when DNA is damaged or when DNA replication is inhibited (Jaffé and D'Ari, 1985). The essential (heterodimeric) DNA synthesis enzyme ribonucleotide reductase (RNR) contains iron (Chitambar and Narasimhan, 1991; Elledge *et al.*, 1992), and indeed when iron is limiting in the cell, DNA synthesis is blocked through inhibition of RNR (Chitambar and Narasimhan, 1991). It has further been shown that blocking DNA replication by inhibition of RNR in *E. coli* induces the SOS response (Barbé *et al.*, 1987). The cell division protein peptidoglycan synthase, encoded by *ftsI*, is an enzyme that is required for synthesis of peptidoglycan in the cell division septum in *E. coli* (Miller *et al*, 2004), and it has been reported that inactivation of FtsI also induces the SOS response (Miller *et al.*, 2004). The gene *nrdB* which encodes the iron-containing (β)- subunit of RNR is between 1.1- and 1.4-fold up-regulated in *F. novicida* at each of the time points sampled, suggesting that this enzyme subunit may be depleted by iron-starvation. It is possible to hypothesise that under iron-depleted conditions DNA replication in *F. novicida* is inhibited through loss of RNR and the SOS response is therefore invoked through down-regulation of FtsI, an event which heralds the reduction in the rate of cell division.

5a.3.2. The *F. tularensis* siderophore was up-regulated

Recent studies by two groups to characterise the response of *F. tularensis* LVS to growth under iron-depleted conditions have revealed an operon that encodes a siderophore (Sullivan *et al.*, 2006; Deng *et al.*, 2006). In *F. tularensis* the siderophore is encoded by genes FTT0026-0029, with the corresponding genes in *F. novicida* encoded by FTN_1682, FTN_1683, FTN_1685, and FTN_1530 (*lysA*). As anticipated, each of these genes was more than two-fold up-regulated by *F. novicida* in response to iron-depletion at each of the three sample time points, however the associated *p*-value was not ≤ 0.05 for every measurement (although the *p*-value for 11 out of the 12 measurements was ≤ 0.1) (table 5a.4). The siderophore is comprised of two genes annotated as from the drug H^+ antiporter-1 (DHA1)- family, A hypothetical protein denoted FrgA, and a diaminopimelate decarboxylase encoded by *lysA*. The two DHA1-family proteins are presumably involved in active transport of iron across the cell wall. The *frgA* gene is annotated as a hypothetical protein, and it is noted that this gene has been shown to be both essential for *F. tularensis* LVS infection of mouse lungs and also under the

Table 5a.4 Fold change of the *F. tularensis* siderophore biosynthesis genes in *F. novicida* cultured with and without iron.

			Fold change		
Gene	**FTN no.**	**FTT no.**	**60 min**	**120 min**	**210 min**
FTN_1685	1685	0026	+7.55 (*p=0.039*)	+7.32 (*p=0.056*)	+10.51 (*p=0.036*)
lysA	1530	0027	+4.91 (*p=0.051*)	+8.24 (*p=0.046*)	+7.19 (*p=0.037*)
FTN_1683	1683	0028	+2.67 (*p=0.234*)	+4.94 (*p=0.081*)	+3.82 (*p=0.040*)
frgA	1682	0029	+7.07 (*p=0.066*)	+7.68 (*p=0.088*)	+9.60 (*p=0.037*)

control of the virulence regulator protein PrfA (Su *et al.*, 2007; Mohapatra *et al.*, 2007). The diaminopimelate decarboxylase encoded by *lysA* catalyzes the conversion of meso-diaminopimelic acid to L-lysine, which is the final step in lysine biosynthesis. A role for *lysA* in bacterial iron acquisition has previously been demonstrated in that LysA is essential for the viability of *M. tuberculosis in vivo* because lysine is required for assembly of mycobactins, a family a membrane-associated siderophores (Krithika *et al.*, 2006). All of the *F. tularensis* siderophore genes except FTT0026 (FTN_1685) were reported to be required for growth and survival of *F. novicida* in murine spleens after infection via the intra-peritoneal (i.p.) route (Weiss *et al.*, 2007).

5a.3.2.1. Further similarities and contrasts between this study and that published by Deng *et al.*

A study published by Deng *et al.*, (2006) set out to characterise the iron-starvation response of *F. tularensis* LVS using sodium dodecyl sulphate-polyacrylamide gel electrophoresis to identify envelope proteins that were up-regulated and a DNA microarray to examine global transcriptome regulation in response to iron-starvation. Including the siderophore biosynthesis genes discussed above (5a.4.2.), the DNA microarray experiment revealed approximately 80 CDS that were at least two-fold regulated in either direction. In both this and the Deng *et al.* study, transport genes from the major facilitator superfamily (MFS) were up-regulated, and chaperones were down-regulated. However, in spite of these similarities, it was noted that (apart from the siderophore biosynthesis genes) no other regulated CDS were observed in common between the Deng *et al.* study and the study

presented here. It is possible that differences in experimental design between these two iron-starvation response studies may account for some of the differences in data obtained. In the *F. tularensis* LVS study bacteria were cultured in Mueller-Hinton media, whereas in this (*F. novicida*) study bacteria were cultured in CDM. A key difference between the experimental designs of the two studies is that Deng *et al.* failed to use an RNA stabilisation reagent such as RNA*later* (ABI) before harvesting bacteria by centrifugation for RNA isolation. RNA stabilisation reagents work by preventing RNase-mediated degradation of RNA, and by preventing further transcription, to provide a 'snapshot' of the transcriptome at the time of sampling. RNA stabilisation is particularly important when performing transcriptomics studies using bacteria, as the half-life of an average RNA can be measured in minutes, allowing bacteria to respond immediately to environmental changes (Sarkar, 1997; Taljanidisz *et al.*, 1997). The failure of Deng *et al.* to use an RNA stabilising reagent means it is likely that the transcriptome response of *F. tularensis* LVS to laboratory manipulation was captured, in addition to the response to the iron-replete or iron-deficient conditions of interest. Both the test and control cultures were exposed to centrifugal force, meaning that, due to the comparative nature of microarray experiments, no genes would be observed as regulated in response to this spurious condition. However, any commonality in the response to centrifugal force and the response to iron-starvation would also be missed, resulting in false-negative results. Differences between the results of the two studies may also, in part, be attributed to differences between *F. novicida* and *F. tularensis* LVS, indeed some key differences between the ways in which the two species respond to

iron-deprivation were reported in the Deng *et al.* paper. It was reported that it was necessary to pre-culture *F. tularensis* LVS in iron-deficient growth medium overnight before sub-culturing in iron-deficient medium in order to observe an altered growth rate between iron-replete and iron-deficient media, whereas this was not a requirement for *F. novicida*, and is supported by the data presented in this study. It was also reported that it was not possible to use the first gene in the *F. novicida* siderophore operon, FTN_1685, to complement a *F. tularensis* LVS mutant for the corresponding gene. Additionally, a *F. novicida* mutant for FTN_1685 was slightly repressed for growth in macrophages and markedly reduced in siderophore biosynthesis whereas the corresponding mutant in *F. tularensis* LVS was not repressed for growth in macrophages and only showed a minor reduction in siderophore biosynthesis. These data indicate that differences do exist between the ways in which *F. tularensis* LVS and *F. novicida* store iron and between the ways in which they respond to iron-starvation, which may also account for the contrasting, and yet in some respects broadly similar, data obtained in these two studies.

5a.3.3. A second putative iron-acquisition gene cluster was up-regulated

A second putative gene cluster encoded by *F. novicida* CDS FTN_0753, FTN_754 (*coaD*), and FTN_755 was up-regulated after 120 and 210 minutes iron-depletion, although, again, some of these results may be considered as indications only because the associated *p*-values were > 0.05 (table 5a.5). The CDS encoded by FTN_0753 is annotated as a hypothetical

Table 5a.5 Fold change of a second putative cluster of iron-related genes in *F. novicida* cultured with and without iron.

			Fold change		
Gene	**FTN no.**	**FTT no.**	**60 min**	**120 min**	**210 min**
FTN_0753	0753	0580	+1.23 (p=0.039)	+1.50 (p=0.051)	+1.75 (p=0.335)
coaD	0754	0581	+1.30 (p=0.043)	+1.75 (p=0.351)	+1.45 (p=0.036)
FTN_0755	0755	0582 (*fdx*)	+1.23 (p=0.050)	+1.67 (p=0.035)	+1.81 (p=0.700)

protein, but it does share a common domain with PaaD, a predicated metal-sulphur cluster biosynthetic enzyme (Refseq. accession number NC_008601). The protein encoded by *coaD* is phosphopantetheine adenylyltransferase, which catalyses a step in the synthesis of coenzyme A (CoA) through the transfer of an adenylyl group from ATP to 4'-phosphopantetheine in the presence of magnesium to give 3'-dephospho-CoA and pyrophosphate (Izard, 2002). Although CoA is an essential coenzyme for wide variety of important reactions, it may be that the by-product pyrophosphate is more important in the acquisition of iron as pyrophosphate is widely reported to facilitate the release of ferric iron from the mammalian iron-storage protein transferrin, and also from chicken conalbumin, in a reaction that has a half-life in excess of one billion years without the presence of pyrophosphate (Carver and Frieden, 1978; Morgan, 1979; Konopka and Romslo, 1981; Perrotte-Piquemal *et al.*, 1999). The third gene in the putative cluster is FTN_0755, which is annotated as a [4Fe-4S] ferredoxin, giving a clear role as an iron-related protein. The up-regulation of three CDS that are adjacently encoded in the genome and are all putatively annotated with functions related to iron acquisition and storage strongly suggests that these genes are important in the response of *F. novicida* to low-iron conditions, although the absence of any putative membrane transport proteins makes it is unlikely that these gene products form a classical siderophore.

5a.3.4. Fe^{2+}-related genes were up-regulated

The Fe^{2+} transport protein encoded by *feoB* was up-regulated by *F. novicida* after 60 minutes iron-depletion. The Feo iron-transport system is important for supply of iron to *E. coli* under anaerobic conditions, where iron is in the ferrous rather than the ferric form (Rocha *et al.*, 1996). Fe^{3+} would be captured by the siderophore and perhaps by the product of second gene cluster both discussed previously (sections 5a.4.2. and 5a.4.3.), but it is intuitive that under severe iron depletion *F. novicida* would also attempt to capture Fe^{2+} through the up-regulation of Feo proteins.

Bacterioferritin, encoded by *bfr*, is, like the closely related ferritin, an iron-sequestering protein with a central iron storage cavity (Andrews *et al.*, 1993), and was up-regulated by *F. novicida* after 120 minutes iron-depletion. In *E. coli* the Bfr protein takes up Fe^{2+} and stores it as a hydrated ferric oxide mineral, which becomes oxidised by a ferroxidase centre (Le Brun *et al.*, 1995). Ferritins are usually considered to play a role in iron-storage rather than in iron acquisition, making up-regulation of these proteins by *F. novicida* under conditions of iron-depletion counter-intuitive, however this result is not without precedent: *E. coli* mutants for ferritin and bacterioferritin are impaired for growth in iron-deficient media, and bacterioferritin has been shown to play a role in *V. cholerae* iron acquisition and virulence (Abdul-Tehrani *et al.*, 1999; Mey *et al.*, 2005).

5a.3.5. Iron-sulphur cluster-related genes were up-regulated

Iron-sulphur clusters are susceptible to iron-starvation, so the up-regulation of two genes involved in the repair of these clusters by *F. novicida*

in response to iron-depletion was not surprising: *rubA*, which encodes rubredoxin, and *sufB*, which encodes an activator of the cysteine desulfurase SufS, were both up-regulated after 210 minutes. Rubredoxin is an iron-sulphur protein that mediates electron transfer in metabolic reactions and, like the closely related iron-sulphur protein ferredoxin (which contains two iron atoms, rubredoxin containing only one), has the ability to accept and discharge electrons, the effect being a change in oxidation state of the iron atoms (Bachmayer *et al.*, 1967). It has been proposed that ferredoxin (or in this case rubredoxin) adjusts the redox state of the iron-sulphur cluster either before or after insertion into a protein, indeed *E. coli* ferredoxin (*fdx*) mutants are unable to fully activate iron-sulphur clusters (Djaman *et al*, 2004). It has previously been reported that cysteine desulfurase is required for synthesis and repair of iron-sulphur clusters (Djaman *et al.*, 2004). In *E. coli* two genes encode separate cysteine desulfurases: *iscS*, which is prominent in *de novo* cluster synthesis, and *sufS*, which is though to play a role in repair of clusters that have lost one or more iron atoms (Patzer and Hantke, 1999). Cysteine desulfurases release sulphur from cysteines for donation to iron-sulphur groups, so the exact mechanism by which this enzyme repairs clusters that have become deficient for iron is unclear, however it has previously been demonstrated that genes of the *suf* operon are strongly induced under conditions of limited iron, as well as under other stress conditions (Outten *et al.*, 2004; Loiseau *et al.*, 2003). As well as activating *sufS*, it has been proposed that SufB also serves to accept sulphur that is liberated by SufS and to act a site for cluster assembly (Layer *et al.*, 2007). In summary, it seems likely that *F. novicida* repairs iron-sulphur clusters that have become

depleted for iron by up-regulating *sufB*, which may serve both to activate the cluster repair enzyme SufS and also as a site for cluster assembly, and by up-regulating rubredoxin, encoded by *rubA*, which donates an electron to clusters which have become positively charged through loss of an iron ion.

5a.3.6. Genes involved in pyrimidine biosynthesis were up-regulated

Three genes from the pyrimidine metabolism pathway, *pyrD*, *pyrF*, and *dnaN*, (fig. 5a.8) were up-regulated by *F. novicida* in response to iron starvation. Diroorotate dehydrogenase is an iron-sulphur cluster containing flavoenzyme encoded by *pyrD*, and orotidine-5'-phosphate decarboxylase, encoded by *pyrF*, are both involved in the *de novo* synthesis of pyrimidines, 95% of which are used as precursors of nucleic acids (Ghim and Neuhard, 1994). DnaN forms the β-chain of DNA polymerase III, which is the chromosomal replicase, and has been shown in *E. coli* to be part of the SOS response, induced by exposure to UV light and by stationary phase growth (Villarroya *et al.*, 1998). A fourth pyrimidine-related gene, FTN_0362, that encodes a deoxyribodipyrimidine photolyase-related protein (also known as DNA photolyase) was also up-regulated by *F. novicida* in response to iron-deficient growth conditions. DNA-photolyase, a light activated flavoenzyme that contains an iron-sulphur cluster, binds pyrimidine dimers in DNA and repairs them to monomers by mediating an electron transfer reaction (Rebeil *et al.*, 1998; DeRosa *et al.*, 2005). It has been shown in *S. enterica* that mutants defective for iron-sulphur cluster metabolism are also auxotrophic for thiamine (Dougherty and Downs, 2006). The up-regulation of genes involved in pyrimidine synthesis and repair could indicate that *F. novicida* detects the

lack of iron in the growth medium and puts effort into synthesising essential pyrimidines using iron reserves that are readily available before these

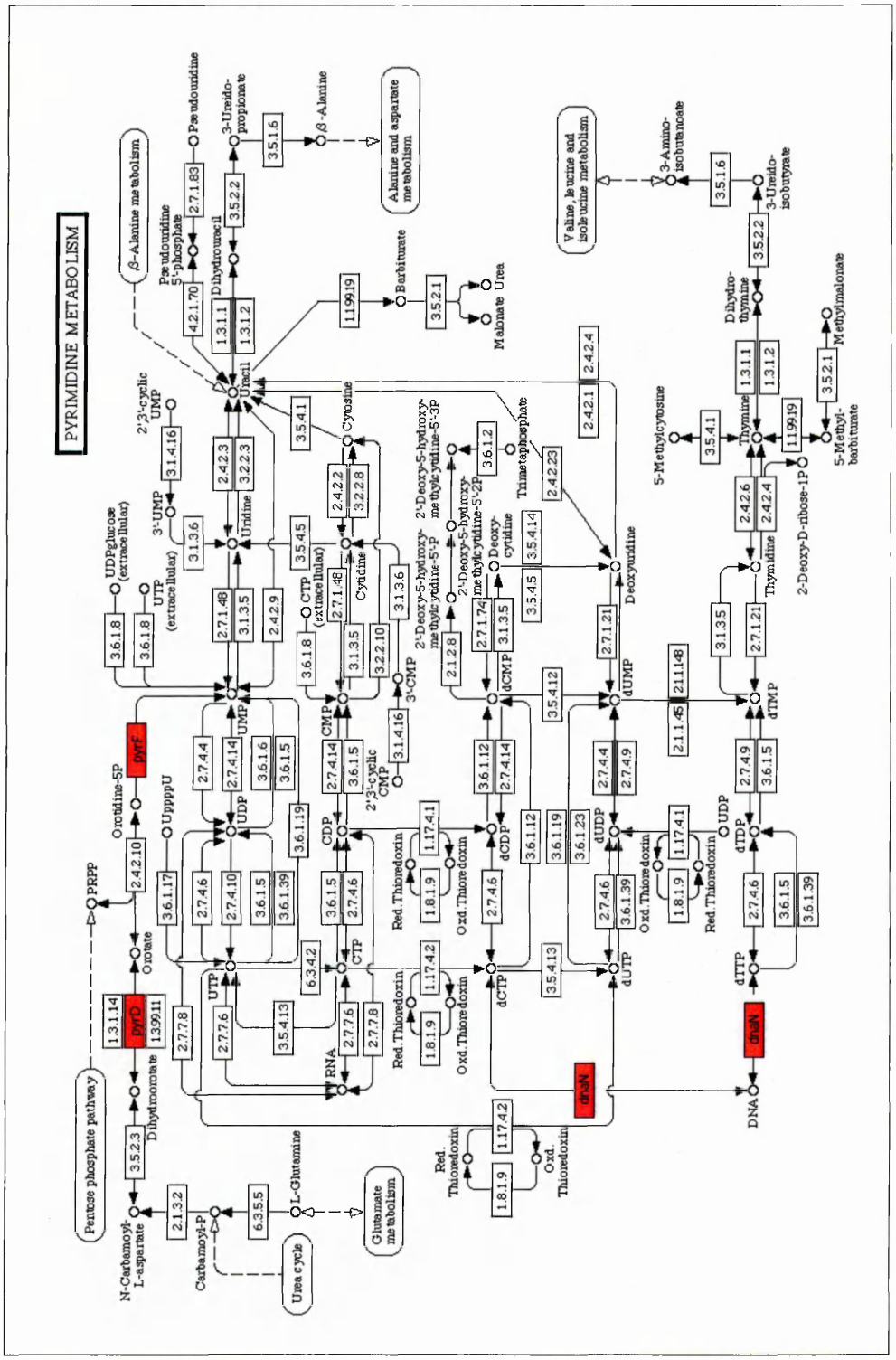


Fig. 5a.8 Pyrimidine metabolism pathway (KEGG, 2007). Genes highlighted in red (*pyrD*, *pyrF*, and *dnaN*) were up-regulated by *F. novicida* in response to iron-starvation.

become fully depleted. Genes involved in the biosynthesis of purines and pyrimidines have been previously identified as virulence factors in a number of bacterial species, for example *Pasturella multocida* (Boyce *et al.*, 2002), and several *pyr* genes, including *pyrD* and *pyrF*, were shown to be required for growth and survival of *F. novicida* in murine spleens after infection via the i.p. route (Weiss *et al.*, 2007).

5a.3.7. Genes that play a role in infection were down-regulated

A number of genes with a demonstrated role in *Francisella* infection were down-regulated in response to iron-depletion: FTT1238 which encodes a membrane protein, *hisS* which encodes a histidyl-tRNA synthetase, *kdpD* which encodes the histidine kinase of an orphaned two-component response regulator, were all reported to be required for murine infection by *F. novicida* via the i.p. route (Weiss *et al.*, 2007). The ATP-dependent protease ClpP is widely reported to function as a HSP in *E. coli*, contributing to the proteolysis of improperly folded or abnormal proteins that tend to occur more frequently under conditions of environmental stress (Kitagawa *et al.*, 1991; Squires *et al.*, 1991). ClpP is both assisted and regulated by a family of Clp ATPases, including ClpB. It has been previously reported that *clpB* is required by both *Rhodococcus equi* for infection of equine macrophages (Rahman *et al.*, 2005), by *F. tularensis* LVS during murine infection (Su *et al.*, 2007), and by *F. novicida* during murine infection (Weiss *et al.*, 2007), making down-regulation of this gene by *F. novicida* under ion-starvation conditions surprising. However, ClpB has also been shown to negatively control the *clp* operon in *E. coli* (Martin, 1996), so perhaps a similar mechanism where ClpP

is expressed via down-regulation of a repressor (ClpB) could exist in *F. novicida*. Because of the role in general stress response that has been attributed to the Clp-protease family, it is difficult to say whether *clpB* is down-regulated by *F. novicida* purely in response to iron-depletion, or as part of a more general stress response brought about by the reduction in growth rate.

Another gene with a demonstrated role in *Francisella* infection that was down-regulated by *F. novicida* in response to iron-depleted conditions was *pdpD*. This gene is part of the FPI described by Nano *et al.* (2004) and is annotated as a hypothetical protein in both the *F. novicida* and the *F. tularensis* Schu S4 genomes. A Δ*pdpD* strain of *F. novicida* has been shown to be defective for growth in macrophages and also defective for virulence in mice (Nano *et al.*, 2004). The down-regulation in response to iron-depletion by *F. novicida* of genes that have been shown to play a role in *Francisella* infection could suggest that either *F. novicida* is able to gain access to iron in the intracellular environment, or perhaps that a lack of iron in the intracellular niche is overcome by some other environmental factor(s) to allow expression of these genes.

5a.3.8. LPS O-antigen biosynthesis genes were down-regulated

The down-regulation of two genes involved in the biosynthesis of the O-antigen component of *F. novicida* LPS, *wbtE* and *wbtH*, in response to iron-depletion seems to indicate that *F. novicida* modifies or reduces it's O-antigen in response to this condition. Both of these genes are involved in the

synthesis of 2-acetamido-2-deoxy-D-galacturonamide, which comprises three out of the four sugars found in the *F. novicida* O-antigen repeat unit (Thomas *et al.*, 2007). Expression of LPS has been shown to be required for internalization of *F. tularensis* by macrophages through activation of complement and complement receptors (Löfgren *et al.*, 1983; Fulop *et al*, 1993; Clemens *et al.*, 2005), and several LPS O-antigen genes, including *wbtE* and *wbtH*, have been shown to be required for *F. novicida* infection of mice via the i.p. route (Weiss *et al.*, 2007). A further role for the *F. tularensis* O-antigen in particular is in protection of the bacterium from complement-mediated serum killing, possibly through the masking of antigenic targets (Sandström *et al.*, 1988). These data would seem to indicate an important extracellular role for the O-antigen, and it may be that *F. tularensis* uses a low-iron environment as a signal that it has reached the intracellular niche and that the O-antigen is no longer required, at least not in the same conformation as that required for the extracellular environment. It is noted that growth of *L. pneumophila* is reduced in murine macrophages that have been activated by (*L. pneumophila*) LPS, and that this is achieved, at least in part, through a down-shift in expression of transferrin receptor by the macrophage and hence a reduction in the availability of intracellular iron for bacteria to sequester (Gebran *et al.*, 1995). Activated macrophages are also difficult to infect by *Francisella* strains of all but the highest virulence (Hartley, 2003), and a similar macrophagic protective mechanism could explain why it is in response to iron-deficiency in particular that it would be advantageous to *F. novicida* if the O-antigen was either reduced in amount or changed in conformation.

5a.3.9. Genes encoding components of an efflux pump were down-regulated

Drug resistance in Gram-negative bacteria can be conferred by resistance-nodulation-cell-division (RND) efflux pumps, which are comprised of three essential components: AcrAB, and TolC (Fernandez-Recio *et al.*, 2004). AcrB is an inner membrane drug:proton antiporter i.e. a pump, and TolC is an outer membrane channel or pore. Together these two require the activity of a periplasmic adaptor protein, AcrA, which binds the substrate to be exported and facilitates removal via the pore (Fernandez-Recio *et al.*, 2004). In response to iron-depletion, both *tolC* and FTN_1609, which is annotated as having a protein domain in common with AcrA, were down-regulated by *F. novicida*. The gene encoded by FTN_1610, which is annotated as having a protein domain in common with AcrB, was also potentially down-regulated but not with a satisfactory associated p-value. Components of RND efflux pumps have been shown to export a range of chemically diverse compounds out of the cell, including the *E. coli* siderophore enterobactin (Bleuel *et al.*, 2005), and the *Pseudomonas aeruginosa* major siderophore pyoverdine (Poole *et al.*, 1993; Li *et al.*, 1995). As discussed previously, *F. novicida* up-regulates genes for siderophore biosynthesis in response to iron-depletion (section 5a.4.2.), but this result would suggest that siderophore secretion may actually be reduced or delayed, perhaps *in lieu* of the functional siderophores that the bacteria concurrently synthesise.

5a.3.10. A potential iron insertion protein was down-regulated

The predicted CDS FTN_1665 is annotated as magnesium chelatase, a key enzyme in the biosynthesis of chlorophyll and bacteriochlorophyll, and the predicted CDS in *F. novicida* shares domains in common with both magnesium chelatase and with a predicted ATPase with chaperone activity (Brittnacher *et al.*, 2006). Magnesium chelatase works by utilising the free energy released upon ATP hydrolysis to catalyze the insertion of a magnesium ion into protoporphyrin IX (Reid and Hunter, 2002), and is found only in those organisms which synthesise either chlorophyll or bacteriochlorophyll (Kannangara *et al.*, 1997), a group which does not include *Francisella*. Initially it would seem that a gross error may have occurred in the annotation of the *F. novicida* genome sequence, however, the regulation of this gene by *F. novicida* in response to iron-depletion, combined with the similarity of the predicted protein to a metal ion insertion enzyme and also to an ATPase may indicate that this enzyme is involved in iron ion insertion in *F. novicida*, potentially through ATP-hydrolysis.

5a.3.11. Several ribonucleases were down-regulated

Three genes encoding ribonucleases were down-regulated by *F. novicida* in response to iron-depletion: *orn*, encoding an oligoribonuclease with 3'-5' exoribonuclease activity (Ghosh and Deutscher, 1999); *rnc*, encoding ribonuclease III; and *rnr*, encoding ribonuclease H. Oligoribonucleotides can inhibit gene expression though hybridisation to (complementary) single-stranded DNA which is formed upon transcription initiation, and they therefore can be considered as regulators of expression (Milne *et al.*, 2001). It is therefore possible that, by either hydrolysing or not

hydrolysing oligoribonucleotides, Orn also plays a role in gene regulation and that down-regulation of this oligoribonuclease by *F. novicida* in response to iron-depletion leads to the subsequent down-regulation of other genes. Gene regulation through RNA in response to iron concentration is not without precedent: *ryhB* is a small antisense regulatory RNA that is repressed by Fur in *E. coli*. At least six mRNA transcripts that encode iron-relevant proteins are negatively regulated by *ryhB*, and removal of iron from the cell causes *ryhB* levels rise causing degradation of the target transcripts (Massé *et al.*, 2003). Compared to other bacterial species, very few classical gene expression regulation systems have been identified in *Francisella* species (Mohapatra *et al.*, 2007), and it is possible to speculate that differential regulation of *orn* is one way in which *F. novicida* overcomes this potential limitation. However, the corresponding down-regulation of the two other ribonucleases, *rnc* and *rnr*, both of which primarily target ribosomal RNA (rRNA) (Wang and Bechhofer,., 1997; Cheng *et al.*, 2002), together with the down-regulation of *orn*, would seem to indicate a general decrease in protein expression that is presumably linked to the slower growth rate of *F. novicida* cultured under conditions of iron-depletion, and that a reduction of growth rate and protein turnover in response to any condition would elicit a similar response with respect to these genes.

Chapter 5b

The oxidative stress response of *F. novicida*

5b.1. Introduction

5b.1.1. Oxidative stress

ROS and reactive nitrogen species (RNS) are important mediators of the bactericidal host response. They are capable of damaging biomolecules including proteins, membrane lipids, carbohydrates and nucleotide bases (Imlay, 2008). In phagocytic cells ROS and RNS result from increased oxygen consumption known as the oxidative burst that occurs during ingestion. Entry of *F. tularensis* LVS into interferon-gamma (IFNγ) -activated macrophages initiates the oxidative burst (Fortier *et al.*, 1992), and indeed Lindgren *et al.* showed that mice, and in particular murine macrophages, deficient for production of ROS and RNS are extremely susceptible to *F. tularensis* LVS infection, demonstrating that ROS and RNS play a role in the control of *F. tularensis* infection (Lindgren *et al.*, 2004; Lindgren *et al.*, 2005).

5b.1.1.1. Generation of ROS and RNS

The oxidative burst initiates a signal cascade ultimately activating an NADPH-dependent oxidase complex which initiates a sequence of univalent reductions of oxygen (Clark, 1990). ROS produced include H_2O_2, hydroxyl radicals ($HO^{\cdot}$), and superoxide anions ($O_2^{\cdot -}$) (fig. 5b.1), each having different reaction rates with neighbouring molecules. The ways in which ROS can mediate damage to bacteria are described in section 5b.1.1.2. Less reactive ROS, for example $O_2^{\cdot -}$, are able to diffuse away from the point of origin before exerting their toxic effects which allows them to access

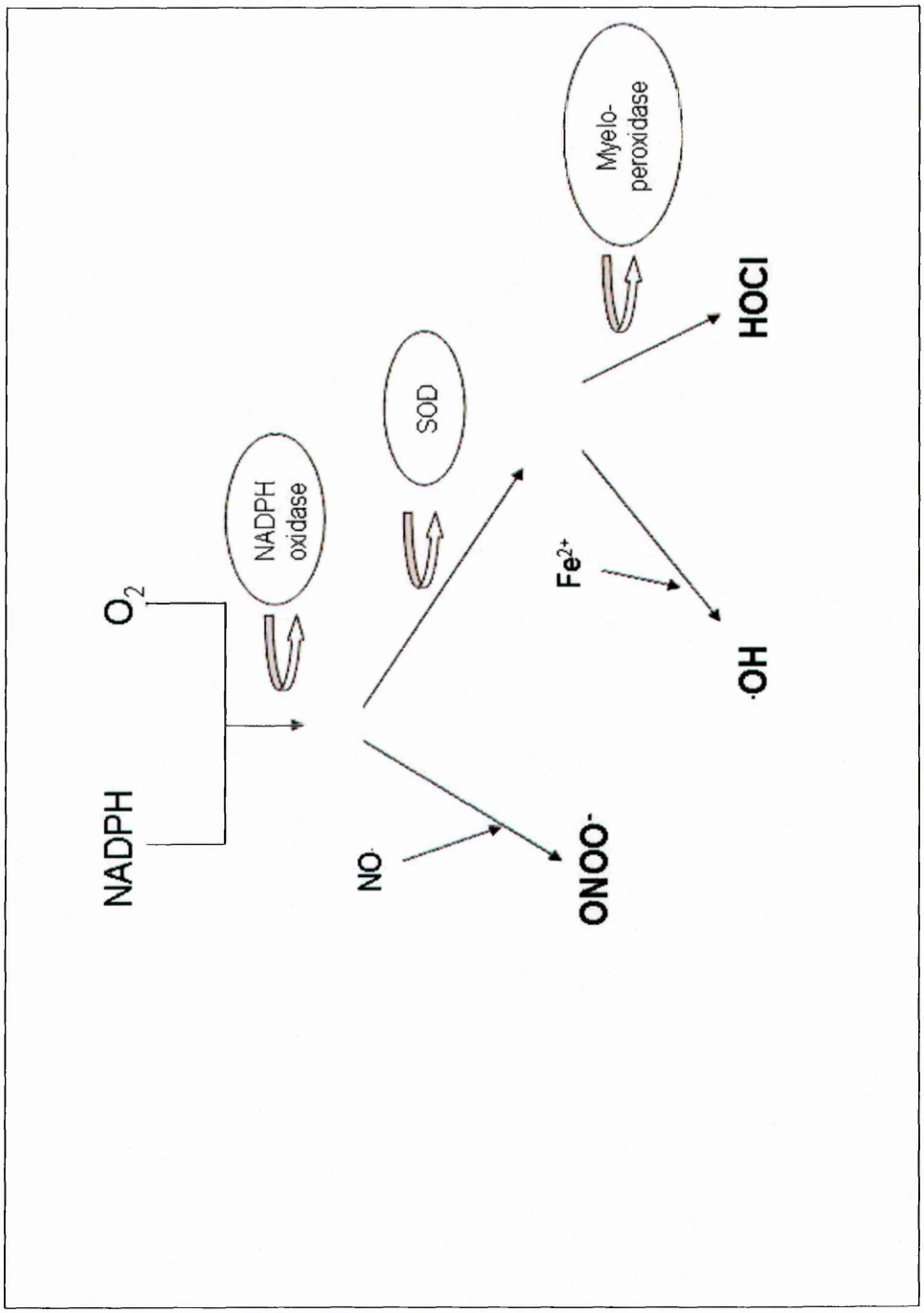


Fig. 5b.1 Schematic of ways in which ROS are produced in the macrophage

substrates inside bacteria, whereas the more reactive ROS, the most highly reactive of which are $HO^{\cdot}$, immediately oxidise the first suitable substrate, usually within a few nanometres of ROS formation (Farr and Kagoma, 1991). This means that damage distal to the site of production of highly reactive ROS occurs via a free radical cascade where the oxidation of one molecule produces a free radical that quickly reacts with its neighbour, setting up a chain reaction.

H_2O_2, which is produced by the self-reaction of $O_2^{\cdot -}$ in the presence of superoxide dismutase (SOD) and is not a free radical (Keyer *et al.*, 1995; McCormick *et al.*, 1998), is more reactive than $O_2^{\cdot -}$ and is therefore readily able to diffuse across membranes. As well as being able to directly oxidize membranes, proteins, and DNA, H_2O_2 is one of the most significant ROS because it participates in the Fenton reaction, in which it reacts with a transition metal (often Fe^{2+}), to yield highly damaging $HO^{\cdot}$ (Keyer *et al.*, 1995; McCormick *et al.*, 1998).

In addition to ROS, RNS are also formed in macrophages and PMNs as a by-product of nitric oxide synthase (NOS) -mediated oxidation of L-arginine and L-citrulline (Nathan and Xie, 1994). Formation of these very potent cytotoxic agents is stimulated by host cytokines including IFNs and tumour necrosis factors (TNFs). Nitric oxide (NO) is an effective oxidant in its own right, or it can react with $O_2^{\cdot -}$ to generate the highly reactive peroxynitrite ($ONOO^-$) and ultimately $HO^{\cdot}$ (Beckman *et al.*, 1990; Miller and Britigan, 1997).

5b.1.1.2. Oxidative damage to macromolecules

Oxidative damage to DNA occurs at base and sugar groups, producing single- and double-strand breaks in the backbone, as well as cross-links to other molecules and lesions that block replication (Cabiscol *et al.*, 2000). Guanine, in particular, can be oxidised to form 8-oxyguanine (a GO lesion) which mis-pairs with adenine so that a mutation is introduced by the substitution of an AT pair in place of a GC pair. The proteins MutM, MutY, and MutT, which together comprise the GO system, are used by *E. coli* to protect itself from GO lesions. MutM and MutY are both base-excision repair glycosylases and MutT is a phosphatase which prevents incorporation of 8-oxyguanine into DNA (Michaels and Miller, 1992). The genome sequence of *F. tularensis* Schu S4 has been annotated to include both *mutM* (FTT0693) and two *mutT* genes (FTT0823 and FTT0984) as well as a third gene from the *mutT*/nudix family (FTT0305), but as yet no *mutY* gene has been identified (Larsson *et al.*, 2005).

ROS and RNIs can also cause damage to both proteins and lipids, having profound effects on the function and integrity of the cell. Oxidation of proteins can occur in several ways as there are many different reactive sites, e.g. the reversible oxidation of sulfhydryl groups to disulfides, adduction of amino acid residues, and protein-protein cross-linking (Fucci *et al.*, 1983). Proteins which contain metal-binding sites are particularly at risk because oxidising agents interact readily with the metal ion itself (Stadtman 1990). Polyunsaturated lipids are major components of cell membranes, and are readily oxidised by both $HO^{\cdot}$ and by $ONOO^{-}$ in self-propagating chain

reactions which can cause serious impairment of the membrane function (Beckman *et al.*, 1990).

5b.1.1.3. Bacterial defence against ROS and RNS

As there is no mechanism for deactivating $HO^{\cdot}$, bacterial defences rely upon reducing the amount of $O_2^{\cdot-}$ and H_2O_2 available for the production of these highly damaging radicals. Characterised bacterial defences against oxidative stress include production of SOD, which converts $O_2^{\cdot-}$ to H_2O_2; production of catalase, which converts H_2O_2 to water and oxygen; and production of endonuclease IV, which excises oxidatively damaged DNA. In *E. coli*, expression of these proteins and others in response to oxidative stress is regulated by the SoxRS and OxyR transcriptional regulators. SoxRS is activated in response to both $O_2^{\cdot-}$ and NO and its regulon includes genes which encode SOD and endonuclease IV (Li *et al.*, 1994; Lynch and Lin, 1996). OxyR is activated in response to H_2O_2 and regulates, amongst others, a gene encoding catalase (Morgan *et al.*, 1986; Imlay and Linn, 1986). The genome of *F. tularensis* Schu S4 contains genes which encode SOD (*sodB* encoding FeSOD and *sodC* encoding Cu,ZnSOD), catalase (*katG*), and the oxidative stress transcriptional regulator *oxyR*. Moreover, a *sodB* mutant of *F. tularensis* LVS (with reduced activity of FeSOD) has been shown to be hypersensitive to H_2O_2 and to $O_2^{\cdot-}$ generated by paraquat (Bakshi *et al.*, 2006). A role for *F. tularensis* catalase (encoded by *katG*) in neutralizing H_2O_2 *in vitro* has also been demonstrated, although a role for this gene in the virulence of *F. tularensis* Schu S4 in mice was not established (Lindgren *et al.*, 2007). At the time of writing, no *F. tularensis* Schu S4 genes

have been annotated as encoding an endonuclease IV or any oxidative stress transcriptional regulators apart from *oxyR* (Larsson *et al.*, 2005).

5b.1.2. Aim

The aim of this study was to determine the pattern of *Francisella* gene expression as a specific response to oxidative stress, a condition that is likely to be encountered by *Francisella* in the macrophage. The *F. tularensis* Schu S4 microarray was used to examine the *F. novicida* response to oxidative stress.

5b.2. Results

5b.2.1. Growth of *F. novicida* under oxidative stress

In order to select biologically relevant sampling times for transcriptome studies, a comparison of the growth rate of *F. novicida* cultured with and without 5 mM H_2O_2 was made. It was observed that *F. novicida* was sensitive to H_2O_2 after one hour (fig. 5b.2), but a growth curve constructed from readings taken within the first hour shows that this sensitivity was not induced until at least 30 minutes after the introduction of H_2O_2 (fig. 5b.3).

5b.2.2. The transcriptome response of *F. novicida* to oxidative stress

Due to the sensitivity of *F. novicida* to 5 mM H_2O_2 at one hour (fig. 5b.3), the transcriptome response to oxidative stress was investigated at two time points: 15 min. and 30 min. As described in section 5.1., *F. novicida* was cultured in CDM-met for transcriptomics analysis in response to oxidative stress. RNA was isolated for each condition on three separate occasions (biological replicates) and each RNA sample was hybridised to the microarray twice (technical replicates). A list of genes that were regulated by *F. novicida* cultured with 5 mM H_2O_2 (compared to without H_2O_2) can be found in table 5b.1.

5b.2.3. Some regulated oligonucleotides were not complementary to *F. novicida* CDS

As mentioned in section 5a.3.2.1., the microarray oligonucleotides used to carry out *F. novicida in vitro* stress response studies were originally designed to be complementary to the *F. tularensis* Schu S4 genome

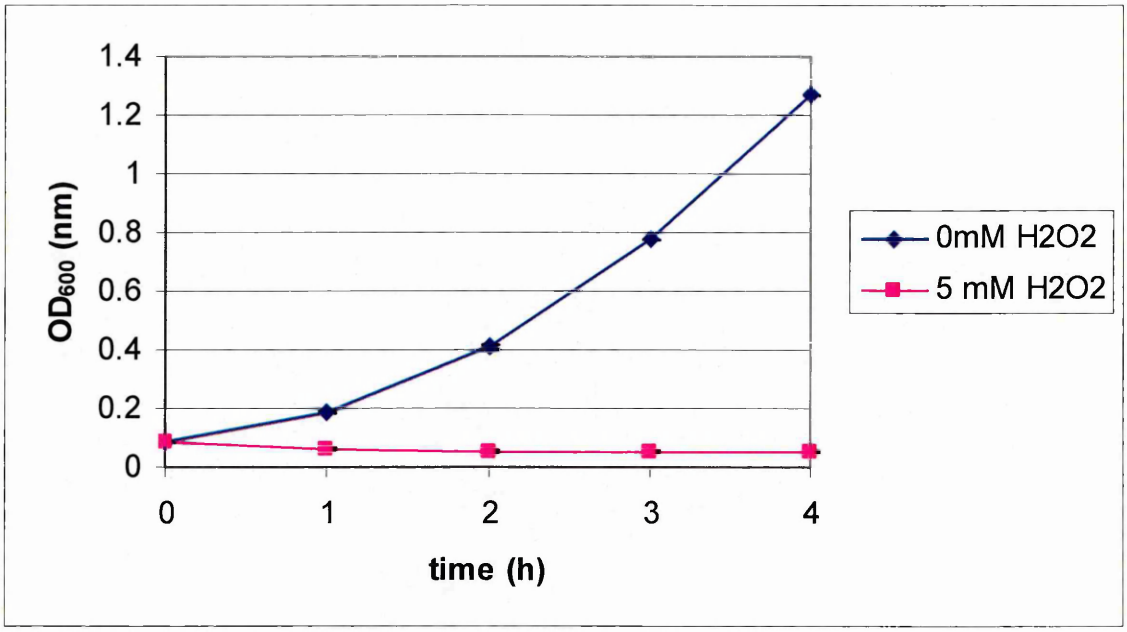


Fig. 5b.2 Comparison of the replication rate of *F. novicida* cultured in CDM with 5 mM H_2O_2 vs. without H_2O_2. Within the first hour after addition of H_2O_2 the growth of the test culture had ceased.

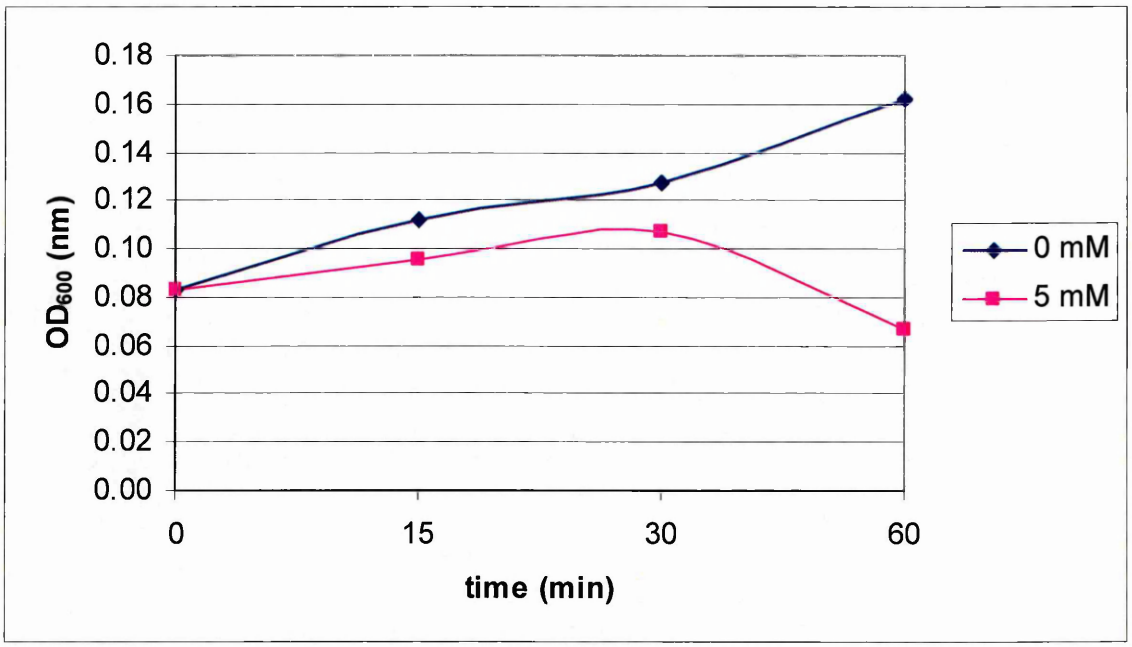


Fig. 5b.3 Comparison of the replication rate of *F. novicida* cultured in CDM with 5 mM H_2O_2 vs. without H_2O_2 over one hour. It was not until 30 min. after addition of H_2O_2 that the growth of *F. novicida* ceased.

Table 5b.1 Genes regulated by *F. novicida* in response to culture under oxidative stress conditions compared to without oxidative stress. [a]Locus tag assigned in the *F. novicida* U112 genome sequence (Brittnacher *et al.*, 2006). [b]Locus tag assigned in the *F. tularensis* Schu S4 genome sequence (Larsson *et al.*, 2005). [c]Pseudogene in *F. tularensis* Schu S4. [d]The nucleotide sequence of FTN_0952 is homologous to the intergenic region between FTT1062 and FTT1063 (*hemF*) in *F. tularensis* Schu S4.

ID	FTN no.[a]	*FTT no.*[b]	Time	Fold change	P_{FDR}	Annotation	GenProtEC classification
Up							
ppdK	0064	*FTT0250*	15	8.31	0.04	Phosphoenolpyruvate synthase/pyruvate phosphate dikinase	[12]
FTN_0312	0312	*FTT1618*[c]	15	6.74	0.05	Drug H+ antiporter-1 (DHA1) family protein	[4.2.A]
glyQ	0519	*FTT0419*	15	3.32	0.05	Glycyl-tRNA synthetase alpha chain	[2.3.1] [7.1]
FTN_0990	0990	*FTT0516*[c]	15	2.74	0.04	4Fe-4S Ferredoxin, FAD-dependent	[12]
ptsN	1295	*FTT1280*	15	7.02	0.05	PEP-dependent sugar phosphotransferase system (PTS), enzyme IIA	[4.9.B]
rnfB	1034	*FTT0649*	15	2.77	0.04	Iron-sulfur cluster-binding protein	[1.4.3]
thrS	1191	*FTT1817*	15	1.55	0.04	Threonyl-tRNA synthetase	[2.3.1] [7.1]
FTN_1542	1542	*FTT1693*	15	1.57	0.04	Hypothetical protein	[13]
Down							
FTN_0092	0092	*FTT0219*	15	1.48	0.04	Inorganic phosphate transporter (PiT) family protein	[12]
tolQ	0352	*FTT0837*	15	1.83	0.05	Group A colicin translocation; TolQ protein	[6.1] [7.3] [8.1] [8.4]
FTN_0777	0777	*FTT1097*	15	1.47	0.05	Hypothetical protein	[13]
FTN_0952	0952	*None*[d]	15	1.52	0.05	Hypothetical protein	[13]
iscS	1245	*FTT1226*	15	3.08	0.05	Cysteine desulfarase	[1.5.1] [2.3.3]
lysA	1530	*FTT0027*	15	1.52	0.05	Diaminopimelate decarboxylase	[1.5.1.7]
FTN_1580	1580	*FTT0136*	15	1.50	0.05	DNA helicase	[12]
frgA	1682	*FTT0029*	15	2.51	0.04	Siderophore biosynthesis protein	[11]
rpsA	0159	*FTT0183*	30	1.63	0.05	30S ribosomal protein S1	[2.3.2] [2.3.8] [6.6] [7.1]
FTN_0326	0326	*FTT1610*[c]	30	1.77	0.05	Hypothetical protein	[13]
FTN_0581	0581	*FTT0490*	30	1.55	0.05	Phospholipase D family protein.	[12]
gltB	1360	*FTT1398*[c]	30	41.40	0.03	Glutamate synthase domain 2	[1.5.1] [1.8.3]
uvrD	1594	*FTT0121*	30	3.02	0.05	DNA helicase II	[2.1.1] [2.1.4] [5.8] [7.1]

sequence. With this in mind, the sequence of each oligonucleotide that was identified as regulated by *F. novicida* was compared to the *F. novicida* genome sequence that was subsequently available (Brittnacher *et al.*, 2006) using the BLASTn program (Altschul *et al.*, 1990). In most cases 100% complementarity was achieved between the oligonucleotide and the *F. novicida* genome sequence and within the *F. novicida* CDS corresponding to the *F. tularensis* Schu S4 CDS to which the oligonucleotide was designed. If less than 100% complementarity was achieved between the oligonucleotide and the *F. novicida* genome, then the most likely *F. novicida* binding site was selected according to the criteria of ≥ 75% overall complementarity and ≥ 15 base pairs contiguous complementarity (Kane *et al.*, 2000). In these cases the sequence of the *F. novicida* CDS selected was then used to find the corresponding CDS in *F. tularensis* Schu S4 using tBLASTx (Altschul *et al.*, 1990) to ensure continuity when interpreting the results in the context of either genome. In the case of FTN_0952 (table 5b.1) which is annotated as a hypothetical protein, no homologue of the *F. novicida* CDS could be identified in the *F. tularensis* Schu S4 genome sequence. Additionally, for oligonucleotide 1096 (oligo1096), designed to *F. tularensis* Schu S4 locus FTT1121, which was down-regulated after 15 min. oxidative stress, the only regions of acceptable complementarity in *F. novicida* were at the two separate intergenic regions, 1) between FTN_1738 and FTN_1739 (trpB) and 2) between FTN_676 (*rpmA*) and FTN_677.

5b.2.4. Regulated CDS were grouped according to function

Each of the CDS regulated by *F. novicida* in response to elevated temperature was assigned to a functional category as previously defined in the GenProtEC database (Serres *et al.*, 2004) (Appedix 1), and represented graphically (fig. 5b.4). The GenProtEC classification assigns gene products to groups according to their broad function, and into sub-groups which describe increasingly specific functions at each level. It should be noted that some CDS are assigned to more than one GenProtEC category.

5b.2.4.1. Functional groups of up-regulated CDS

CDS up-regulated by *F. novicida* in response to oxidative stress were spread between Metabolism, Information transfer, Transport, and Hypothetical groups. Only two of the up-regulated CDS were assigned to a group for final location of gene products, both to the cytoplasm. No up-regulated CDS were assigned to groups for Cell structure or Cell processes.

5b.2.4.2. Functional groups of down-regulated CDS

The functional group containing the largest number of down-regulated CDS was Information transfer, including both DNA- and protein-related subgroups, followed by Metabolism, including amino acid synthesis, and nitrogen metabolism. The gene products of three down-regulated CDS were assigned to a cellular location, two to the cytoplasm and one to the inner membrane. Two down-regulated CDS were assigned to Cell Structure sub-categories, the membrane and ribosome. The only down-regulated CDS assigned to a Cellular process category was in linked to the SOS response.

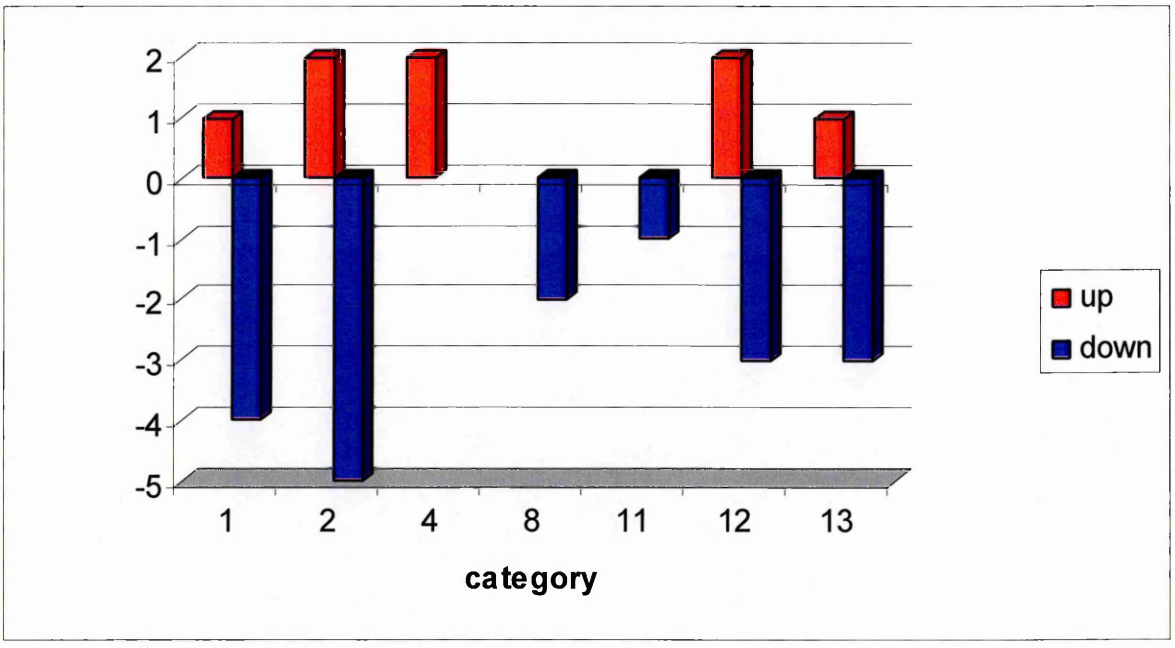


Fig. 5b.4 GenProtEC categories of CDS regulated by *F. novicida* in response to culture under oxidative stress conditions. The x-axis represents GenProtEC categories of CDS regulated in response to oxidative stress were Metabolism (category 1), Information transfer (category 2), Transport (category 4), extrachromosomal genes (category 8), Pathogenic related genes (category 11), Conserved hypothetical genes (category 12), and Hypothetical genes (category 13). The y-axis represents the number of genes up-regulated (+) or down –regulated (-).

5b.2.5. QPCR

As described in section 5a.2.4., TaqMan QPCR was carried out on five CDS that were indicated by microarray as up-regulated by *F. novicida* in response to oxidative stress: FTT0516, FTT1693, *glyQ*, *ppdK*, and *rnfB* (table 5b.1). Assays were performed and the results calculated as described in section 2.7. For each RNA sample, the previously reported *F. tularensis* housekeeping gene, *prfB*, was also assayed to provide a normalisation standard (Nübel et al., 2006). It was noted that the level of expression of *prfB* as determined by microarray was similar in both test and control cultures, supporting the use of this as a housekeeping gene. Microarray results were confirmed by QPCR with a Pearson correlation of 0.945 (fig. 5b.5). As mentioned in section 5a.2.4., the overall Pearson correlation between microarray data and TaqMan QPCR data for 18 genes assayed, selected because they were indicated by microarray as regulated by *F. novicida* in response to one of the conditions tested, was 0.533.

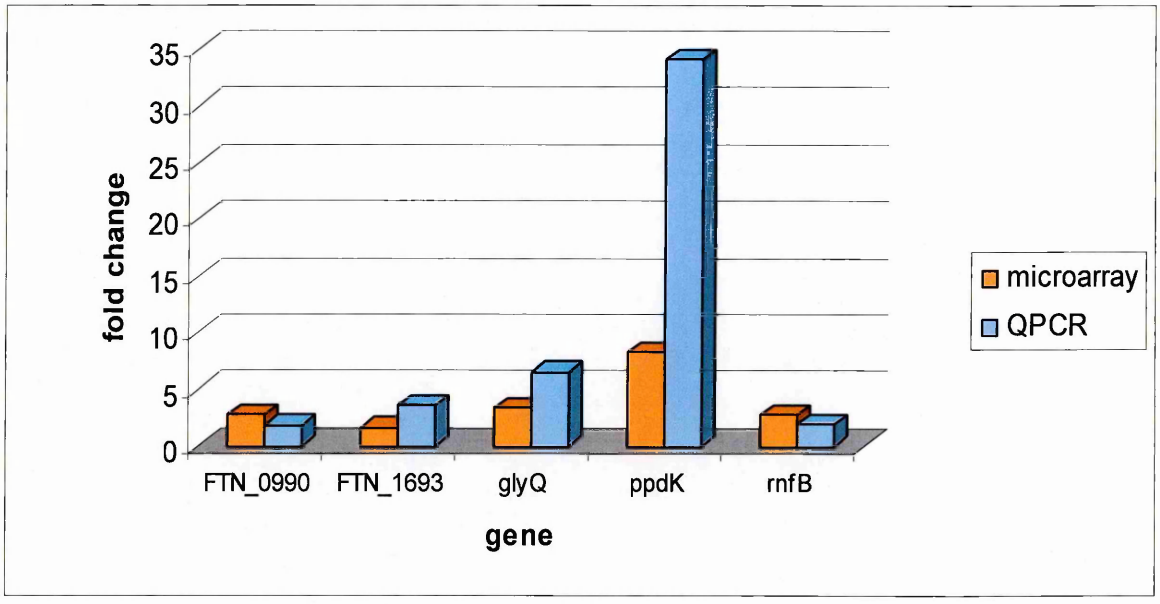


Fig. 5b.5 QPCR confirmation of microarray data. The QPCR data supported the microarray data in that all genes were regulated in the same direction, although not always to the same magnitude:

Gene	Microarray	QPCR
FTN_0990	2.74	1.79
FTN_1693	1.57	3.53
glyQ	3.32	6.57
ppdK	8.31	34.14
rnfB	2.77	1.87
Correlation	**0.94492531**	

5b.3. Discussion

With the exceptions of hypothetical and conserved hypothetical proteins, all of the CDS up-regulated by *F. novicida* in response to oxidative stress fell into one of three GenProtEC classification groups: Metabolism, Transport, or Information transfer. The up-regulated Metabolism gene was an iron-sulphur cluster binding protein, which may be involved in synthesis of iron-sulphur clusters that would be likely to suffer damage under conditions of oxidative stress. A sugar phosphotransferase gene assigned to the GenProtEC Transport category was up-regulated, potentially leading to increased production of pyruvate, which is a scavenger of H_2O_2. Information transfer systems for the production of amino-acyl tRNA synthetases were also up-regulated, possibly acting as a cellular warning system indicating a source of oxidative stress.

CDS down-regulated by *F. novicida* in response to oxidative stress also fell into the GenProtEC categories of Metabolism or Information transfer. Down-regulated CDS assigned to both of these groups were linked to siderophore biosynthesis and also to iron-sulphur cluster repair. Information transfer genes that were down-regulated were linked to DNA replication, which is hypothesised to be down-regulated as part of the reduction in bacterial viability observed in *F. novicida* exposed to oxidative stress.

5b.3.1. No microarray data were obtained for SOD or catalase genes from *F. novicida* under oxidative stress

Previous studies have indicated a role for SOD (encoded by *sodB*) and catalase (encoded by *katG*) in the protection of *F. tularensis* LVS and Schu S4 against H_2O_2 *in vitro* (Bakshi *et al.*, 2006; Lindgren *et al.*, 2007). Unfortunately the probes for *sodB* and *katG* failed to hybridise in this *F. novicida* study, meaning that no data were obtained for either of these genes under oxidative stress conditions. The QPCR studies that were carried out in this study were designed to confirm or refute the microarray data, and as no data were obtained for these two genes, they were not included in QPCR studies in this instance. However, further studies of the oxidative stress response of *F. novicida* would almost certainly include investigation of the behaviour of these two genes in particular.

5b.3.2. Pyruvate synthesis genes were up-regulated

Phosphoenolpyruvate (PEP) synthase, encoded by *ppdK*, was up-regulated by *F. novicida* after 15 minutes exposure to H_2O_2. This enzyme catalyzes the synthesis of PEP from pyruvate, and in turn, PEP is the substrate for pyruvate kinase (encoded by *pyk*) in further pyruvate synthesis (fig. 5b.6). A PEP-dependent sugar phosphotransferase (PTS), encoded by *ptsN*, is also up-regulated by *F. novicida* after 15 minutes exposure to H_2O_2. The PTS catalyses the transport and phosphorylation of a sugar as it passes across the plasma membrane into the bacterial cytoplasm. The PTS-catalyzed reaction yields both a sugar-phosphate and pyruvate (Saier Jr., 1977). Pyruvate is a known scavenger of H_2O_2, producing acetic acid, water,

and carbon dioxide via a non-enzymatic oxidative decarboxylation reaction. Pyruvate has been shown to protect cultured bovine cells from peroxide-induced apoptosis (Kang *et al.*, 2001) and the up-regulation of genes that are involved in pyruvate recycling and synthesis, *ppdK and ptsN*, would seem to be important in the response of *F. novicida* to oxidative stress.

5b.3.3. Genes encoding potential "alarmones" were up-regulated

The gene encoding the α-chain of glycyl-tRNA synthetase, *glyQ*, has been shown by aCGH to be absent or highly divergent in some *F. tularensis* subsp. *holarctica* strains tested, including LVS, (section 4.3.4.4.). However this gene, and also *thrS*, which encodes threonyl-tRNA synthetase, was up-regulated by *F. novicida* after 15 minutes oxidative stress. Aminoacyl-tRNA synthetases attach the correct amino acid to a specific tRNA. They are the interface between the nucleotide sequence and the amino acid sequence and it is therefore intuitive that their function is crucial for accurate gene translation, a cellular function which would be under pressure in conditions of oxidative stress. Interestingly, in both *E. coli* and *S. enterica* serovar Typhimurium, oxidative stress has been linked to the increased synthesis of adenylated nucleotides by aminoacyl-tRNA synthetases, with the hypothesis that the adenylated nucleotides serve to signal the onset of oxidative stress (Lee *et al.*, 1983, Bochner *et al.*, 1984). Additionally, an *E. coli* mutant defective for asparaginyl-tRNA synthetase has been shown to be temperature sensitive (Yamamoto *et al.*, 1977) supporting the hypothesis that adenylated nucleotides are important in bacterial stress response. These data would seem to support the previous studies, in particular the idea that

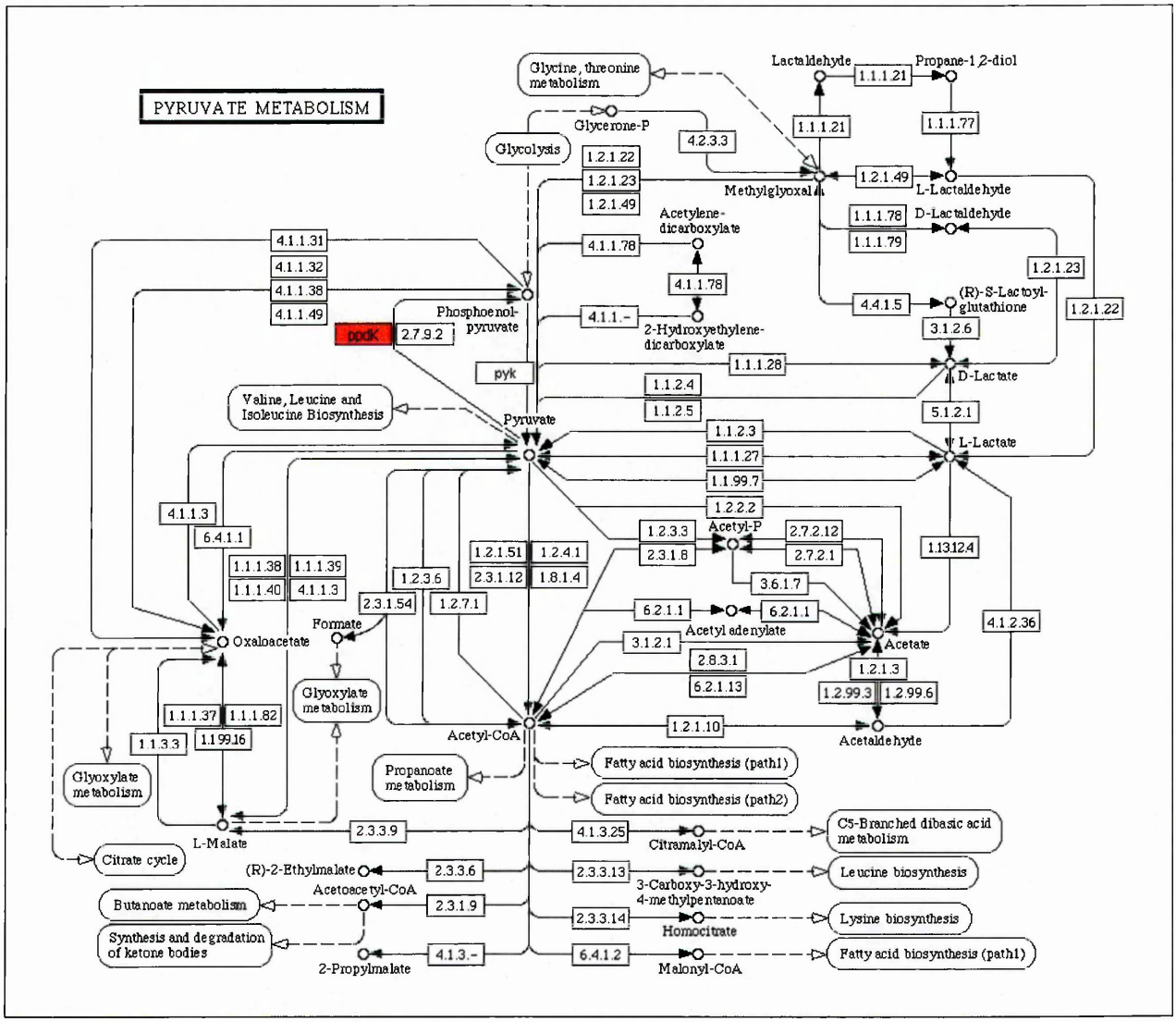


Fig. 5b.6 Pyruvate biosynthesis pathway (KEGG, 2007). The gene highlighted in red (*ppdK*) was up-regulated by *F. novicida* in response to oxidative stress.

adenylated nucleotides may provide a warning system, in that this gene is expressed transiently and towards the beginning of a period of oxidative stress. The absence of this gene in some strains of *F. tularensis* subsp. *holarctica* may indicate one of the virulence factors for which this relatively attenuated subsp. is deficient.

5b.3. 4. Iron-sulphur cluster genes were regulated

Ferredoxins are iron-sulphur proteins that can be thought of as 'biological capacitors' in that they act as electron transfer agents in biological redox reactions so that the oxidation state of the iron atoms changes between 2^+ and 3^+ (Cammack *et al.*, 1971). Bacterial ferredoxins, which contain a cluster of four iron and four sulphur atoms, are susceptible to H2O2, with the loss of an electron destabilising the cluster causing the release of one iron atom thereby removing the catalytic properties of the cluster (Rosen and Klebanoff, 1985; Djaman *et al.*, 2004). Release of iron from iron-sulphur clusters may also cause further damage to the bacterial cell via the Fenton reaction, in which the reduced iron reacts with H_2O_2 to produce highly reactive, and therefore damaging, hydroxyl radicals (Pianzzola *et al.*, 1996), making it important for bacteria to re-capture the iron by prompt repair of iron-sulphur proteins. It may be in response to the loss of functional ferredoxin that *F. novicida* up-regulates FTN_0990, an CDS thought to encode the same, which confirms previous studies in which iron-sulphur proteins, including ferredoxins, have been shown to be up-regulated in response to oxidative stress (Jung *et al.*, 1999; van Vliet *et al.*, 2001; Alamuri *et al.*, 2006). A functionally-related gene, *mfB*, was up-regulated by

F. novicida in response to oxidative stress. This gene is annotated as an iron-sulphur cluster binding protein which suggests that it may serve to repair or to protect iron-sulphur clusters, or simply for the synthesis of *de novo* ferredoxin. In addition, two siderophore biosynthesis genes, *frgA and lysA* (section 5a.4.2.), were down-regulated by *F. novicida* after 15 minutes exposure to H_2O_2, possibly in response to the oxidative damage-mediated release of iron from iron-sulphur clusters. However, the down-regulation of a cysteine desulfarase, encoded by *iscS*, is difficult to explain as it has previously been demonstrated that *iscS* is involved in repair of iron-sulphur clusters that have become deficient for one iron atom, as well as in the *de novo* synthesis of clusters (Djaman *et al.*, 2004). This enzyme releases sulphur from L-cysteine which is donated to iron-sulphur clusters (Zheng *et al.*, 1998; Alamuri *et al.*, 2006), making the participation of an enzyme that catalyses sulphur donation to repair of clusters missing iron, in itself, counter-intuitive, and may represent the reason that *iscS* is down-regulated in *F. novicida* under (oxidative) conditions where iron is under attack.

5b.3.5. DNA helicases were down-regulated

Two DNA helicases were down-regulated by *F. novicida* under oxidative stress, one encoded by FTN_1580 and the other encoded by *uvrD*. The *uvrD* gene product catalyses the ATP-dependent unwinding of DNA and is associated with the correction of mismatched bases and DNA repair in *E. coli* and *S. enterica* serovar Typhimurium (Pang and Walker, 1983). It is induced during the *E. coli* SOS response and is thought to be regulated by transcription attenuation (Easton and Kushner, 1983). It is surprising that this

gene is down-regulated by *F. novicida* under oxidative stress, as intuitively the opposite ought to be true since H_2O_2 is known to damage DNA. One explanation for this could be that by 30 min. oxidative stress, *F. novicida* genes are regulated in preparation for the loss of viability that was observed after 60 minutes oxidative stress.

5b.3.6. A membrane-associated protein was down-regulated

The inner membrane lipoprotein encoded by *tolQ* was also down-regulated by *F. novicida* in response to oxidative stress. This lipoprotein is widely reported to be essential for outer membrane stability, and is required for uptake of most group A colicins and ss-DNA from some filamentous phage (Lewin and Webster, 1996; Dubuisson *et al.*, 2005). Initially it is surprising that *F. novicida* would down-regulate a protein that is essential for membrane stability, however it has also been reported that mutations in any of the *tol* genes (*tolQRAB*) in *E. coli* results in a mucoid phenotype, with the suggestion that a separate regulator gene senses alterations in the cell surface through a lack of *tol* gene expression and activates capsule synthesis in order to protect the cell (Clavel *et al.*, 1996; Lazzaroni *et al.*, 1999). Is it possible that *F. novicida* down-regulates *tolQ* in order to stimulate formation of a capsule that would protect it from H_2O_2? At the time of writing, no capsule genes have been identified in *F. novicida*, but perhaps it is still possible that the down-regulation of *tolQ* in response to oxidative stress is a now redundant evolutionary response. An alternative explanation for the down-regulation of *tolQ* in response to H_2O_2 can also be proposed: mutations in *tol* genes have been shown to cause variation in porin content in *E. coli* in

that the amount of OmpC is increased, whilst the levels of porins OmpF and LamB are decreased (Lazdunski *et al.*, 1998). OmpF allows larger molecules to pass across the cell membrane than OmpC, and under conditions of stress OmpF is preferentially lost in order to minimise access of toxic substances (Lugtenberg *et al.*, 1976; Hall and Silhavy, 1981; Cohen *et al.*, 1988; Pratt and Silhavy 1996). It has been reported that the down-regulation of OmpF plays an important role in the response of bacteria to toxins (Delihas and Forst, 2001), and that oxidative stress in particular has been shown to cause a decrease in OmpF in *E. coli*, although the mechanism of regulation is through post-transcriptional regulation by *micF*, an anti-sense RNA that is encoded just upstream of *ompC* (Nunoshiba *et al.*, 1993; Chou *et al.*, 1993). There are no homologues of either *ompC*, *ompF*, or *lamB* genes identified in the *F. novicida* or *F. tularensis* Schu S4 genomes, but again it is possible to hypothesise that *F. novicida* may down-regulate *tolQ* in response to H_2O_2 in order that an, as yet unknown, porin may also be down-regulated, limiting the access of H_2O_2 across the cell membrane.

Chapter 5c

The elevated temperature response of *F. novicida*

5c.1. Introduction

Contrary to its title, the adaptive bacterial response known as the heat shock response is actually triggered by a range of environmental conditions in addition to elevated temperature, for example genes regulated in response to H_2O_2 can also be heat shock proteins (HSPs), iron/Fur-regulated, and transcriptional regulators, as well as involved in oxidative response/repair or of unknown function (Stohl *et al.*, 2005). The classical bacterial heat shock response involves the transient increase in expression of a number of HSPs which mainly function either as molecular chaperones to ensure appropriate folding, translocation and assembly of polypeptide structures, or as proteases to ensure the destruction of inappropriately constructed proteins. The heat shock response is not confined to bacteria; a similar although not identical response is induced both in the host and in the pathogen during infection as both will experience the elevated temperature which is characteristic of a fever. In the host this response helps to maintain cell integrity, and contributes to immune signalling and to pathogen recognition (Murray and Young, 1992). Therefore examining the effect of heat shock can provide information about the adaptive response of pathogens to a generally stressful growth condition, rather than simply their response to higher temperatures and therefore can provide indicators as to how bacteria respond to analogously stressful conditions that may be encountered *in vivo*.

5c.1.1. Aim

As previously stated, the determination of stress responses can provide clues as to the survival and virulence mechanisms that *Francisella* employs

during host infection. The aim of these experiments was to examine the specific response of *Francisella* to a generally stressful condition, as opposed to the response of *Francisella* to elevated temperature *per se*. The *F. tularensis* Schu S4 microarray was used to examine the *F. novicida* response to elevated temperature.

5c.2. Results

5c.2.1. Growth of *F. novicida* at elevated temperature

In order to select biologically relevant sampling times for transcriptome studies, a comparison of the growth rate of *F. novicida* cultured in CDM at 42°C and at 37°C (the control) were compared. The growth rate of *F. novicida* was not affected by elevating the culture temperature from 37°C to 42°C (fig. 5c.1).

5c.2.2. The transcriptome response of *F. novicida* to elevated temperature

Due to the similarity between the growth rate of *F. novicida* at 37°C and at 42°C (fig. 5c.1), the transcriptome response of *F. novicida* to elevated temperature was investigated at three time points selected during the first hour post stress-induction: 15 min., 30 min., and 60 min. As described in section 5.1., *F. novicida* was cultured in CDM-Met for transcriptomics analysis in response to elevated temperature. RNA was isolated for each condition on three separate occasions (biological replicates) and hybridised to the microarray twice technical replicates). A list of genes that were regulated by *F. novicida* at 42°C (compared to 37°C) can be found in table 5c.1. As previously described in section 5a.3.2.1., because the microarray oligonucleotides used to carry out *F. novicida in vitro* stress response studies were originally designed to be complementary to the *F. tularensis* Schu S4 genome sequence, the sequence each oligonucleotide that was identified as regulated by *F. novicida* to elevated temperature was compared to the *F. novicida* genome sequence that was subsequently available (Brittnacher *et*

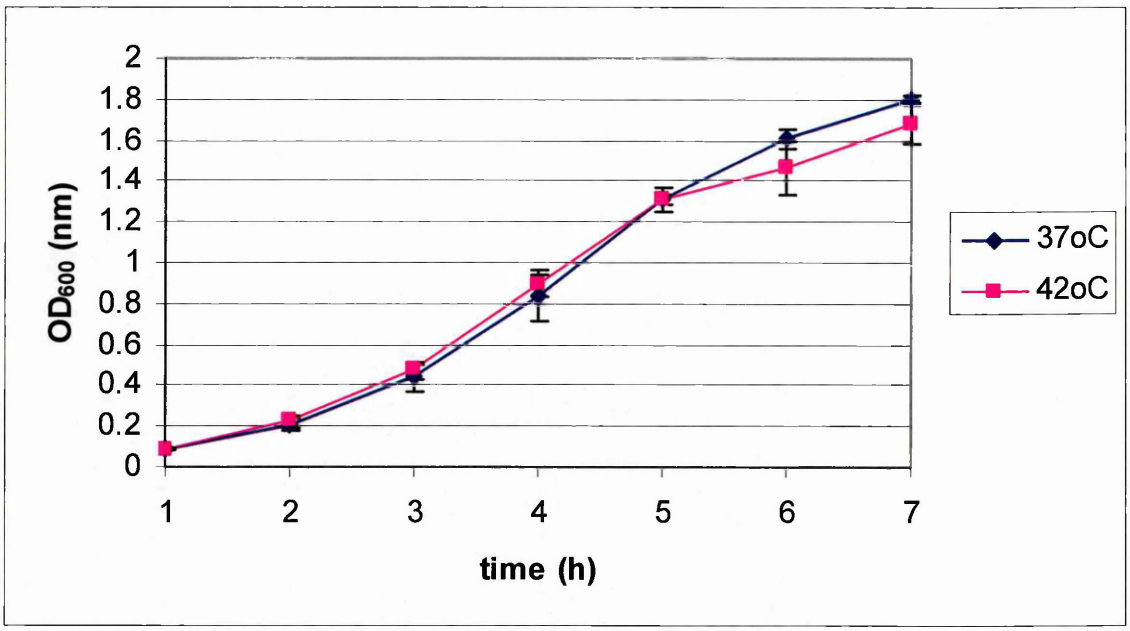


Fig. 5c.1 Comparison of the replication rate of *F. novicida* cultured in CDM at 42°C vs. 37°C. The growth rate was equivalent throughout the assay.

Table 5c.1 Genes up-regulated by *F. novicida* in response to at 42°C compared to 37°C. [a]Locus tag assigned in the *F. novicida* U112 genome sequence (Brittnacher *et al.*, 2006). [b]Locus tag assigned in the *F. tularensis* Schu S4 genome sequence (Larsson *et al.*, 2005). [c]Pseudogene in *F. tularensis* Schu S4.

Time	ID	FTN no.[a]	*FT no.*[b]	Fold change	P_{FDR}	Annotation	GenProtEC Classification
	Up						
15	FTN_1326	1326	*FTT1361*[c]	24.88	0.04	Hypothetical protein	[11] [13]
30	FTN_0859	0859	*FTT0978*	2.90	0.01	Hypothetical membrane protein	[12]
60	FTN_0049	0049	*FTT1068*	4.86	0.05	Hypothetical protein	[13]
60	*galU*	0729	*FTT0757*	2.78	0.05	UTP-glucose-1-phosphate uridylyltransferase	[1.1.1] [1.6.2] [1.7.7] [1.7.9] [5.5.5] [6.7]
60	*dxs*	0896	*FTT1018c*	10.39	0.03	1-deoxy-D-xylulose 5-phosphate synthase	[1.5.3.isoprenoid] [1.5.3.6] [1.5.3.8]
60	FTN_1010	1010	*FTT0672*[c]	4.53	0.05	Major facilitator superfamily (MFS) transport protein	[4.2.A.1]
60	FTN_1053	1053	*FTT0628*	6.28	0.05	Outer membrane protein of unknown function	[7.4]
60	FTN_1068	1068	*FTT0613*	11.39	0.05	Hypothetical protein	[13]
60	*ispF*	1110	*FTT1128*	5.87	0.05	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	[1.5.3.isoprenoid] [7.1]
60	*feoA*	1368	*FTT1403*	5.74	0.05	Fe2+ transport system protein A	[4.9] [4.9.A.8] [5.5.7]
60	FTN_1612	1612	*FTT0103*	6.11	0.05	Hypothetical protein	[13]
60	*sdhC*	1639	*FTT0072*	3.78	0.05	Succinate dehydrogenase, cytochrome b556	[1.3.4] [1.3.6] [1.4.1] [1.4.3] [1.6.15.1] [6.1] [7.3] [11]
60	*nuoA*	1680	*FTT0031*	8.56	0.05	NADH dehydrogenase I, A subunit	[1.3.6] [1.3.7] [1.4.1] [4.3.D.1] [6.1] [7.3]
	Down						
15	FTN_0026	0026	*FTT1658*	3.25	0.05	Hypothetical protein	[13]
15	FTN_0091	0091	*FTT0220*	1.74	0.05	Hypothetical protein	[13]
15	FTN_0923	0923	*FTT1045*	1.69	0.05	Hypothetical protein	[13]
15	FTN_1753	1753	*FTT1759*[c]	1.90	0.05	Rieske (2Fe-2S) domain protein	[12]
30	*slt*	0496	*FTT0400*	4.95	0.02	Soluble lytic murein transglycosylase	[1.6.7] [5.1] [6.2] [7.2] [11]
30	*chiA*	0627	*FTT0715*	7.23	0.05	Chitinase, glycosyl hydrolase family 18	[1.2] [7.2] [11]
30	*proS*	1377	*FTT1412*	6.62	0.02	Prolyl-tRNA synthetase	[2.3.amino acid starvation] [7.1]
60	FTN_0067	0067	*FTT0226*[c]	1.52	0.05	Hypothetical protein	[13]
60	FTN_0119	0119	*FTT1747*	3.63	0.05	Conserved outer membrane protein of unknown function	[7.4]

al., 2006) using the BLASTn program (Altschul *et al.*, 1990). In most cases 100% complementarity was achieved between the oligonucleotide and the *F. novicida* genome sequence and within the *F. novicida* CDS corresponding to the *F. tularensis* Schu S4 CDS to which the oligonucleotide was designed. If less than 100% complementarity between the oligonucleotide and the *F. novicida* genome was achieved, then the most likely *F. novicida* binding site was selected according to the criteria of ≥ 75% overall complementarity and ≥ 15 base pairs contiguous complementarity (Kane *et al.*, 2000). In these cases the sequence of the *F. novicida* CDS selected was then used to find the corresponding CDS in *F. tularensis* Schu S4 using tBLASTx (Altschul *et al.*, 1990) to ensure continuity when interpreting the results in the context of either genome.

5c.2.3. Regulated CDS were grouped according to function

Each of the CDS regulated by *F. novicida* in response to elevated temperature was assigned to a functional category as previously defined in GenProtEC (Serres *et al.*, 2004) (Appendix 1), and represented graphically (fig. 5c.2). The GenProtEC classification assigns gene products to groups according to their broad function, and into sub-groups which describe increasingly specific functions at each level. It should be noted that some CDS are assigned to more than one GenProtEC category.

5c.2.3.1. Functional groups of up-regulated CDS

The functional group containing the highest number of up-regulated CDS was Metabolism (fig. 5c.3). Within the Metabolism functional group, the

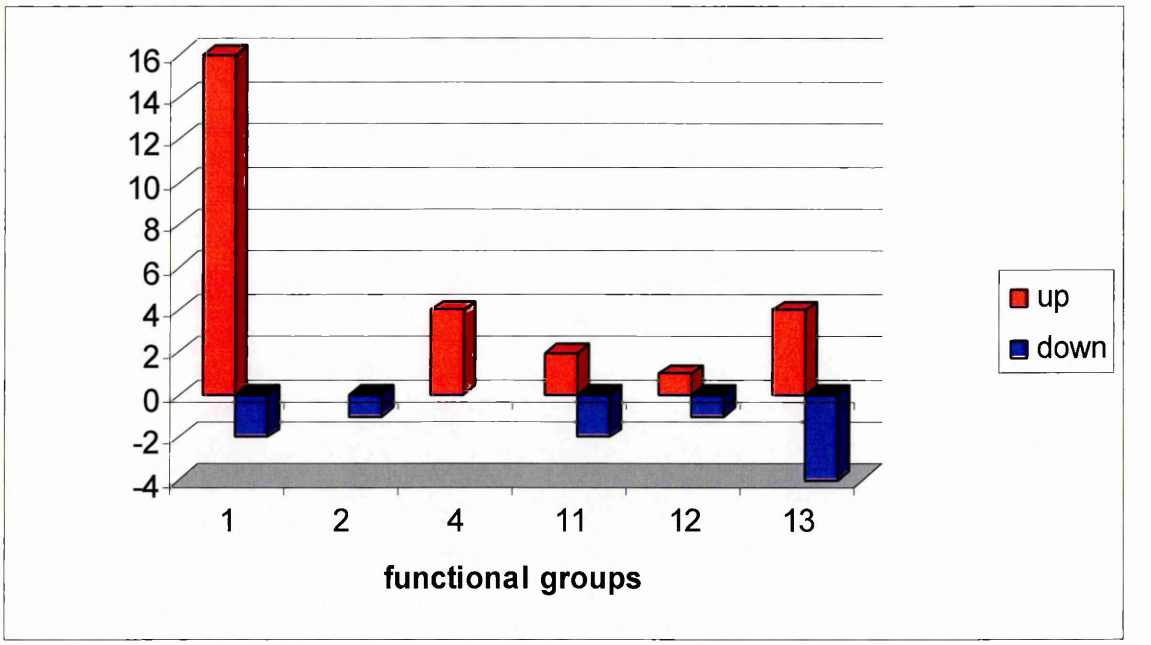


Fig. 5c.2 GenProtEC categories of CDS regulated by *F. novicida* in response to culture under elevated temperature. The x-axis represents GenProtEC categories of CDS regulated in response to elevated temperature were Metabolism (category 1), Information transfer (category 2), Transport (category 4), Pathogenic related genes (category 11), Conserved hypothetical genes (category 12), and Hypothetical genes (category 13). The y-axis represents the number of genes up-regulated (+) or down –regulated (-).

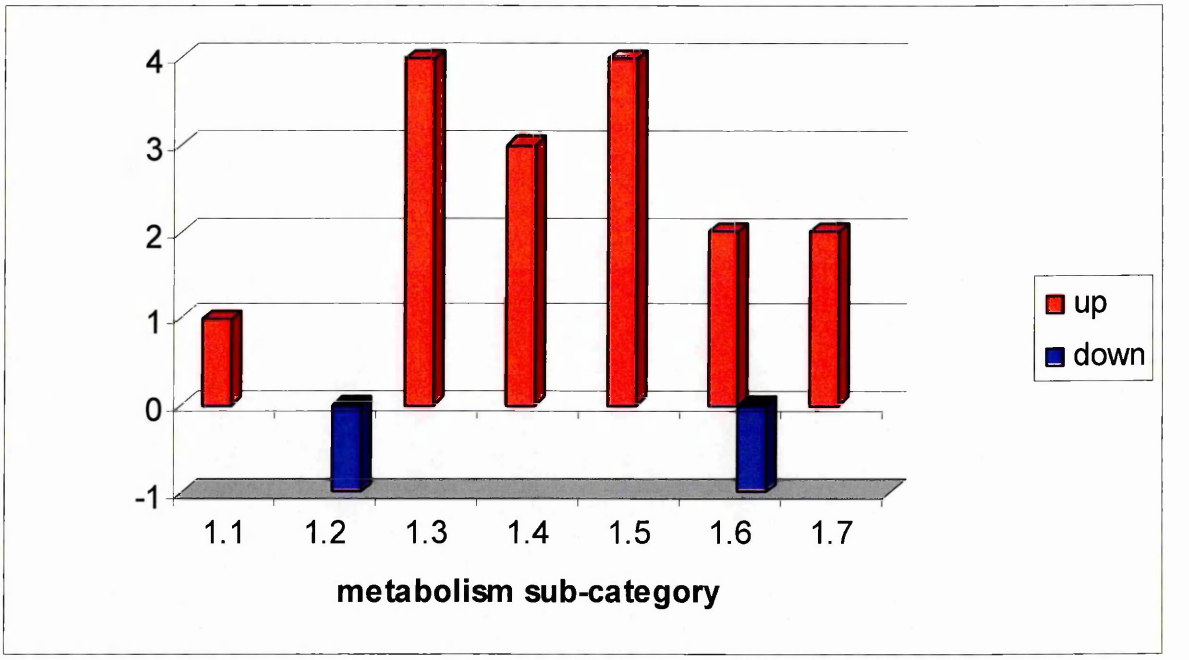


Fig. 5c.3 GenProtEC Metabolism sub-categories of CDS regulated by *F. novicida* in response to culture under elevated temperature. The x-axis represents GenProtEC Metabolism sub-categories of CDS regulated in response to elevated temperature were Carbon compounds metabolism (category 1.1), Macromolecule degradation (category 1.2), Energy metabolism, carbon (category 1.3), Energy production, transport (category 1.4), Building block biosynthesis (category 1.5), Macromolecule biosynthesis (category 1.6), and Central intermediary metabolism (category 1.7). The y-axis represents the number of genes up-regulated (+) or down –regulated (-).

up-regulated CDS were assigned to macromolecular biosynthesis (M-antigen and cytochromes), building block biosynthesis (isoprenoids, vitamin B6, and thiamin), energy production and transport (electron donor and carrier), energy metabolism (aerobic and anaerobic respiration, and tricarboxylic acid cycle), central intermediary metabolism (glucose and galactose metabolism), and carbon compounds metabolism. The second largest functional group of up-regulated CDS was Transport, including ferrous uptake family-, electrochemical potential driven- and primary active-transporters. Of the four up-regulated CDS for which a cellular location of the final gene product was assigned, two were located in the inner membrane, one was located in the outer membrane, and one was located in the cytoplasm (fig. 5c.4). Three up-regulated CDS were assigned to a cell structure sub-category: two were assigned to the Membrane, and one to Capsule. Two up-regulated CDS were assigned to the Cell Processes category: one to Desiccation and one to Iron acquisition.

5c.2.3.2. Functional groups of down-regulated CDS

The functional group containing the largest number of down-regulated CDS was also Hypothetical proteins, followed by Metabolism (Macromolecule biosynthesis and degradation) (fig. 5c.3) and Pathogenic-related genes. Of the four down-regulated CDS for which a cellular location of the final gene product was assigned, two were located in the periplasm, one was located in the outer membrane, and one was located in the cytoplasm (fig. 5c.4). The only down-regulated CDS assigned to a Cell Structure sub-category was

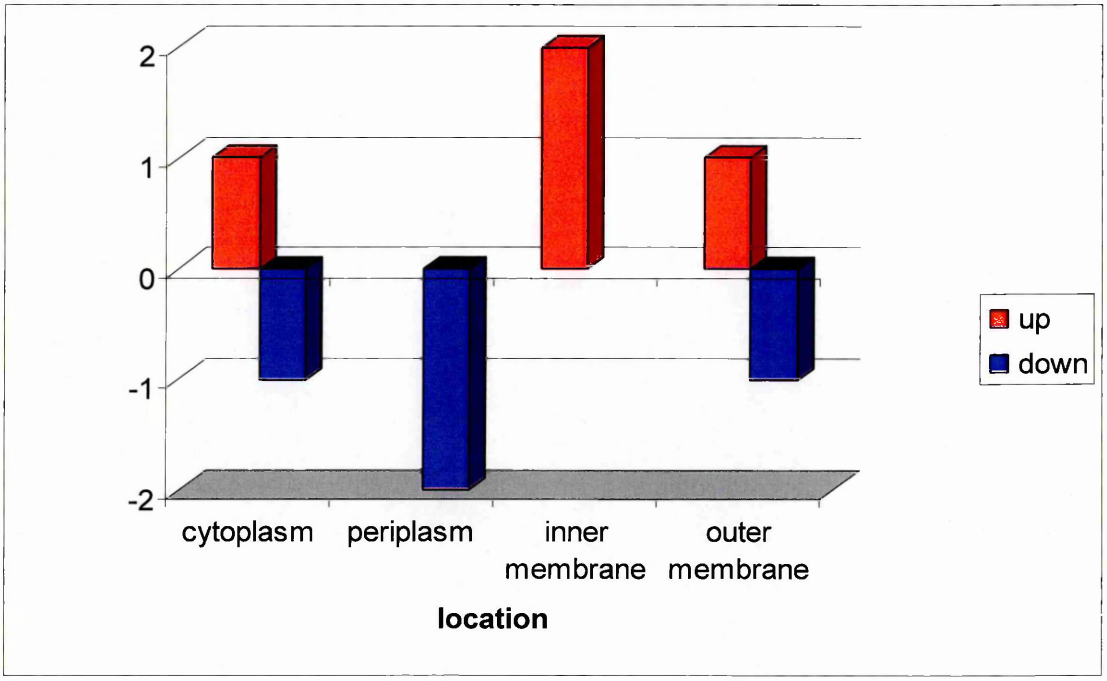


Fig. 5c.4 GenProtEC Cellular locations of CDS regulated by *F. novicida* in response to culture under elevated temperature (x-axis). The y-axis represents the number of genes up-regulated (+) or down –regulated (-).

linked to Peptidoglycan, and the only down-regulated CDS assigned to a Cellular processes category was in linked to cell division.

5c.2.4. QPCR

TaqMan QPCR was carried out on five CDS that were indicated by microarray as regulated by *F. novicida* in response to elevated temperature: FTT1361 and FTT0978 were up-regulated, and FTT1658, FTT0715, and *proS* were down-regulated (table 5c.1). For each RNA sample, the previously reported *F. tularensis* housekeeping gene, *prfB*, was also assayed to provide a normalisation standard (Nübel *et al.*, 2006). It was noted that the level of expression of *prfB* as determined by microarray was similar in both test and control cultures, supporting the use of this as a housekeeping gene. These microarray results were confirmed by QPCR with a Pearson correlation of 0.966 (fig. 5c.5). As mentioned in section 5a.3.4., the overall Pearson correlation between microarray data and TaqMan QPCR data for 18 genes assayed, selected because they were indicated by microarray as regulated by *F. novicida* in response to one of the conditions tested was 0.533.

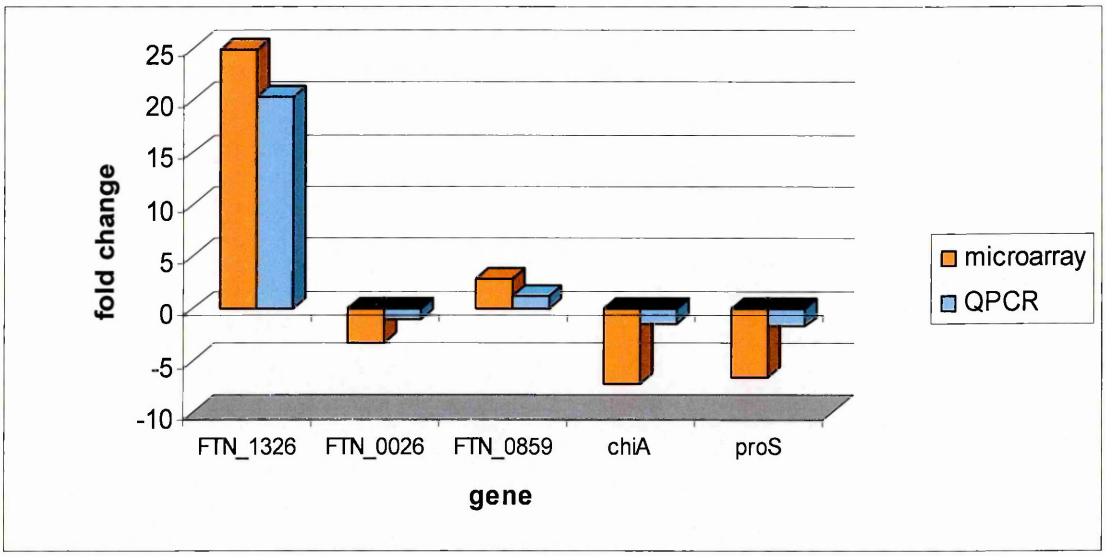


Fig. 5c.5 QPCR confirmation of microarray data. The QPCR data supported the microarray data in that all genes were regulated in the same direction, although not always to the same magnitude:

Gene	Microarray	QPCR
FTN_1326	24.88	20.35
FTN_0026	-3.25	-1
FTN_0859	2.9	1.23
chiA	-7.23	-1.5
proS	-6.62	-1.74
Correlation	**0.9655319**	

5c.3. Discussion

5c.3.1. *Francisella* does react to culture at elevated temperatures

The growth rate of *F. novicida* was not affected by elevating the culture temperature from 37°C to 42°C (fig. 5c.1). This confirms the findings of a previous study carried out by Ericsson *et al.*, who reported that the survival of *F. tularensis* LVS was not affected by culture at 42°C (compared to 37°C), and that the rate of protein synthesis remained stable for at least one hour after the shift in temperature (Ericsson *et al.*, 1994). The similarity between the *Francisella* growth rates at 37°C and 42°C may imply that that the bacteria do not react to the increase in temperature, however, in the study by Ericsson *et al.* mentioned above, the increased expression of a number of proteins, including heat shock proteins (HSPs), was observed (Ericsson *et al.*, 1994), indicating that regulation of gene expression is a requirement for the maintenance of the growth rate by *F. tularensis* at elevated temperatures.

5c.3.2. The transcriptome response of *F. novicida* to elevated temperature

Several groups of CDS that were regulated by *F. novicida* in response to elevated temperature were assigned to the GenProtEC classification groups of Metabolism and/or Pathogen-related genes. Metabolic systems that were up-regulated included genes that were involved steroid biosynthesis, possibly in order to maintain cellular membrane integrity, and genes involved in energy production and metabolism, likely to be required by bacteria in order to cope with the increased physiological burdens imposed

by elevated temperature. Pathogenic-related genes that were up-regulated included a putative HSP. Genes for the metabolism of nucleotide sugars were both up and down-regulated, again possibly in order to reinforce the cell membrane.

5c.3.3. HSPs

HSPs, as the name suggests, are a group of proteins whose rate of expression is transiently elevated by cells (both prokaryotic and eukaryotic) in response to a rise in temperature (Georgopoulos and Ang, 1990). Most HSPs fall into one of two categories: chaperones, the main role of which is to assist in the correct folding of newly synthesised proteins and to prevent their aggregation but without becoming part of the final protein or protein complex themselves, and proteases, which remove improperly folded or otherwise abnormal proteins (Arsène *et al.*, 2000). Both chaperones and proteases are required for growth at all temperatures, but become more important for growth at higher temperatures when the frequency of mis-folding or aggregation becomes higher (Hecker *et al.*, 1996).

The *F. tularensis* Schu S4 and *F. novicida* genomes encode at least 14 known HSPs, including at least six chaperones and two ATP-dependent proteases (Larsson *et al.*, 2005; Brittnacher *et al.*, 2006). A previous proteomics study of the response of *F. tularensis* LVS to a shift in temperature from 37°C to 42°C showed the up-regulation of at least 15 proteins, most prominently the chaperones DnaK, GroEL, and GroES, in samples taken up to 35 minutes after exposure to the higher temperature

(Ericsson *et al.*, 1994). However, in this study only one potential HSP was observed to be up-regulated in *F. novicida* after the temperature shift from 37°C to 42°C: FTN_1326 was up-regulated within 15 minutes but was not observed again, fitting with the assertion that HSPs are transiently expressed. In a previous study FTN_1326 was putatively annotated as *pmcA* (Nano *et al.*, 2004) which is homologous to putative molecular chaperones, in particular the VdcC protein of *B. subtilis* (Nano *et al.*, 2004; Broekhuijsen *et al.*, 2003; Samrakandi *et al.*, 2004). This gene corresponds to CDS FTT1361, a pseudogene in *F. tularensis* Schu S4, and also coincides with $RD_{holarctica}6$ (section 4.3.5.6.) meaning that the CDS is either absent or highly divergent in many *F. tularensis* subsp. *holarctica* strains (Broekhuijsen *et al.*, 2003; Samrakandi *et al.*, 2004). This RD is also coincident with the previously described FPI (Nano *et al.*, 2004). FTN_1326 has also been demonstrated, using an *in vivo* negative selection assay, to be required for growth and survival of *F. novicida* in murine spleens after infection via the i.p. route (Weiss *et al.*, 2007). Since the gene corresponding to FTN_1326 is either missing or non-functional in *F. tularensis* subsp. *tularensis* and subsp. *holarctica*, both of which are virulent to a greater or lesser degree in several mammalian species including humans, it becomes apparent that the intracellular survival of *F. novicida*, and perhaps the pathogenicity that it does display in some hosts, may well be mediated by very different gene expression patterns to those of *F. tularensis*.

Of the remaining known *Francisella* HSPs, which would be expected to be up-regulated upon temperature increase, only five were potentially up-

regulated (chaperones *grpE*, *secB*, and *groES*, and the ATP-dependent protease ClpP subunits *clpP* and *clpX*), and nine were potentially down-regulated, disappointingly none of these had an associated p-value that was appropriate to pass the criteria for designation as regulated in either direction.

5c.3.4. Genes involved in energy release were up-regulated

Two genes involved in oxidative phosphorylation, *nuoA* and *sdhC*, (fig. 5c.6) were up-regulated by *F. novicida* in response to the elevated temperature. Oxidative phosphorylation is the process by which energy, in the form of adenosine 5'-triphosphate (ATP) is stored by the cell. The up-regulation of genes in the oxidative phosphorylation pathway indicates that there is a requirement for additional energy to overcome the physiological stresses imposed by heat shock. Zavilgelsky *et al.* have reported that HSPs require ATP to re-fold thermo-inactivated luciferase expressed in *E. coli* (Zavilgelsky *et al.*, 2002) and as the two main heat shock proteases encoded by *Francisella*, Lon and ClpP, are both ATP-dependent it seems likely that a similar requirement for ATP could exist to drive the heat stress response mechanisms. The *Francisella* aCGH study showed that *sdhC* is absent or highly divergent in some *F. tularensis* subsp. *holarctica* strains tested, including LVS, (section 4.3.4.4.). The gene *sdhC* encodes succinate dehydrogenase, cytochrome B, and was up-regulated by *F. novicida* after 60 minutes growth at the elevated temperature, and has also been shown to be required for growth and survival of *F. novicida* in murine spleens after infection via the i.p. route, with a *F. novicida* Δ*sdhC* strain demonstrated by

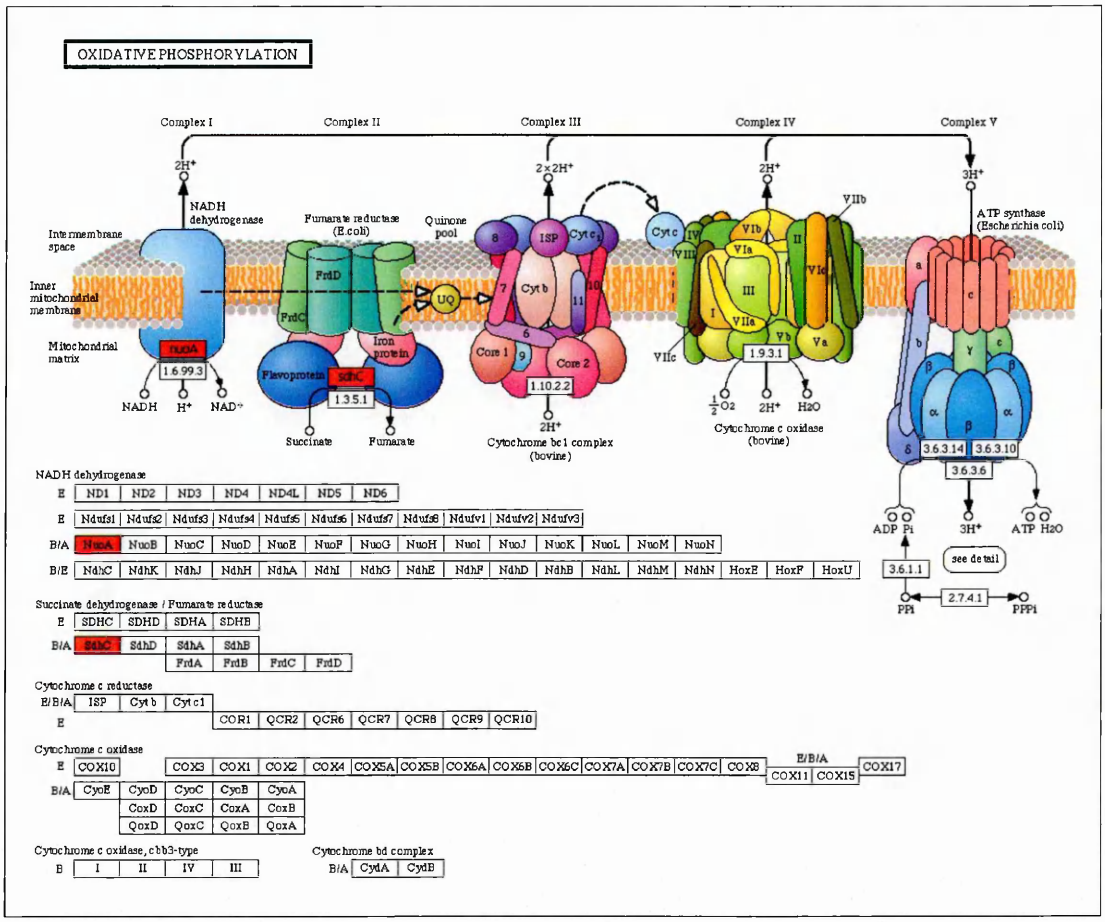


Fig. 5c.6 Oxidative phosphorylation pathway (KEGG, 2007). Genes highlighted in red *(nuoA* and *sdhC*) were up-regulated.

competition index to be mildly attenuated *in vivo* (Weiss *et al.*, 2007). In *E. coli*, cytochrome b556 forms part of complex II which catalyzes the reduction of ubquinone by succinate (Nakamura 1996). In mice, mutants in the SDHC gene of complex II are unable to protect against mitochondrial oxidative stress damage to nuclear DNA (Ishii *et al.*, 2005). Taken together this may indicate that *sdhC* is required for an effective stress response by *Francisella*, and that the absence of this gene in some strains of *F. tularensis* subsp. *holarctica* could be one of the attenuating factors of this subsp.

5c.3.5. Steroid biosynthesis genes were up-regulated

Also up-regulated by *F. novicida* in response to elevated temperature were two genes from the steroid biosynthesis pathway, *dxs* and *ispF* (fig. 5c.7). The products of these two genes are specifically part of the pathway for the formation of isopentenyl-diphosphate which is required for the biosynthesis of all of the steroids shown. Two steroid biosynthetic pathways of particular interest in the context of a bacterial heat shock response are those of terpenoid and porphyrin biosynthesis. Both of these classes of molecules are found in all living cells. Porphyrins combine readily with metal ions e.g. heme is iron II protoporphyrin-IX complex which can then become assimilated into enzymes e.g. heme is found in both oxidase and catalase. It is noted that after 60 min at 42°C the gene encoding the ferrous iron transport protein, *feoA*, is also up-regulated, potentially providing additional iron for complex with porphyrins. Terpenoids play a role in all cellular membranes, in particular sterols, which are complex terpenoids, influence membrane fluidity. It is possible to hypothesise roles for both porphyrins and

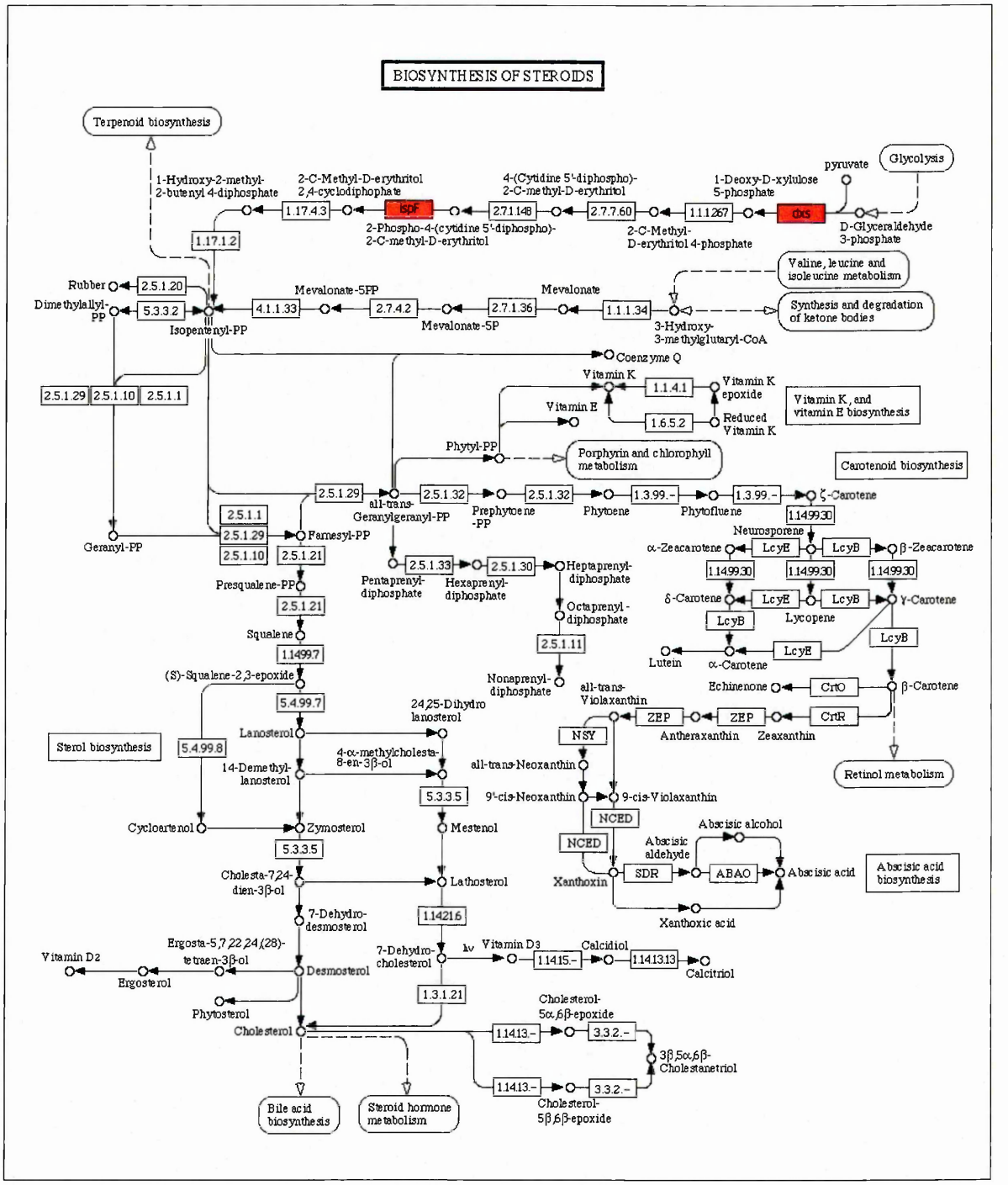


Fig. 5c.7 Steroid biosynthesis pathway (KEGG, 2007). Genes highlighted in red (*ispF* and *dxs*) were up-regulated.

terpenoids in the bacterial heat shock response: porphyrins can provide additional metal binding groups for the enzymes that are likely to be required, and terpenoids can play a role in maintaining membrane integrity which is likely to be adversely affected by heat. Of the remaining five genes in the steroid biosynthesis pathway, only *ispD* was putatively up-regulated, however the p-values for all five genes were higher than the acceptable range to give statistically significant results.

5c.3.6. Genes involved in metabolism of nucleotide sugars were regulated

Two genes that are included in the KEGG pathway describing the metabolism of nucleotide sugars were down-regulated by *F. novicida* in response to elevated temperature, and one gene was up-regulated (fig. 5c.8). The down-regulated genes were *slt*, which putatively encodes a soluble lytic murein transglycosylase, and *chiA*, which putatively encodes a chitinase. The substrate both of these enzymes is described in the pathway as 'pentosans' which simply refers to homopolysaccharides which degrade to pentoses upon hydrolysis. The pentosan substrate for the ChiA protein would appear to be based on arabinose and the substrate for Slt would appear to be based on xylose. Down-regulation of both of these genes would suggest that 'pentosans' are lacking in *F. novicida* cultured at 42°C. Chitinase encoded by *chiA* can be considered as a potential virulence factor for intracellular bacteria since *L. pneumophila* has been shown to express chitinase (*chiA*) upon intra-tracheal inoculation into murine lungs (DebRoy *et al.*, 2006). In *E. coli*, Slt belongs to a superfamily of glyosidases and it is

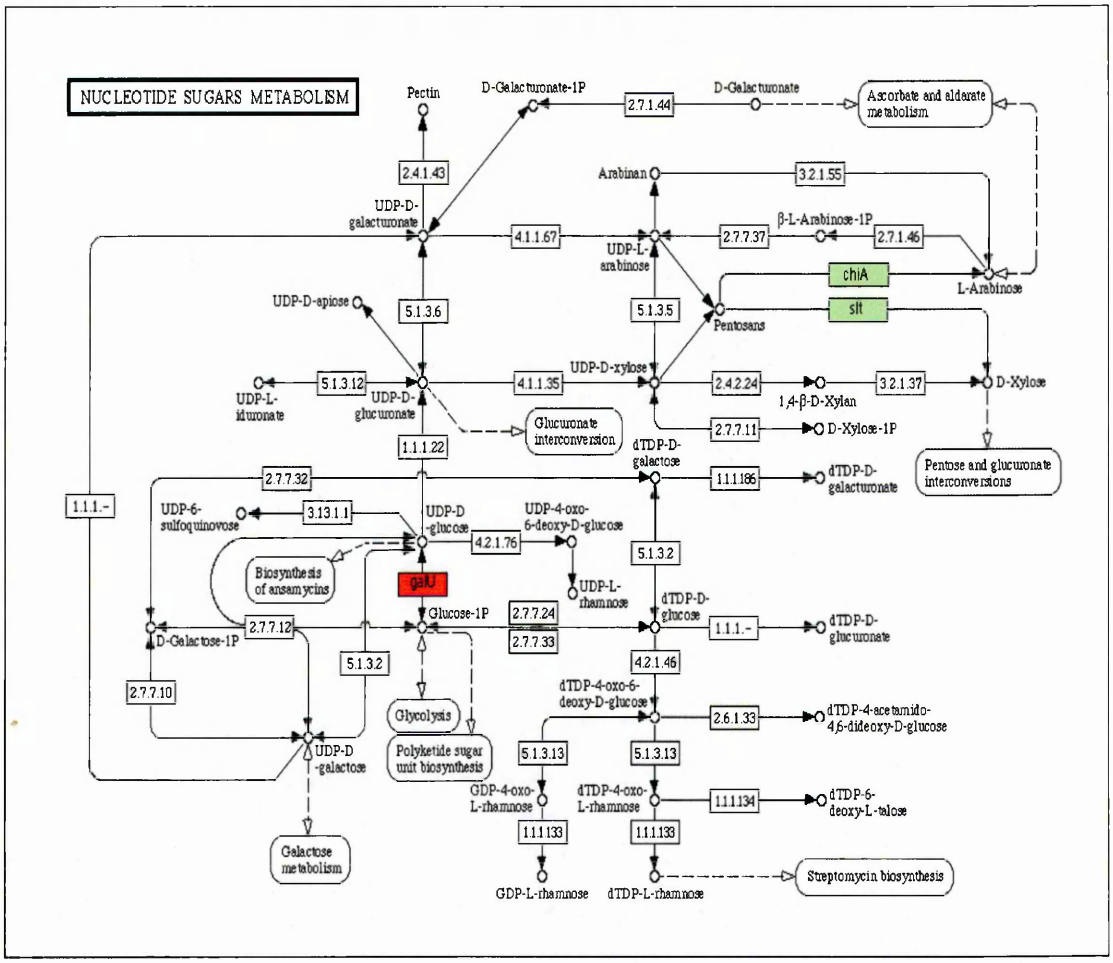


Fig. 5c.8 Nucleotide sugars metabolism pathway (KEGG, 2007). The gene highlighted in red *(galU)* was up-regulated, and the genes highlighted in green (*chiA* and *slt*) were down-regulated.

reported to play a role in correct assembly of the peptidoglycan layer (Chaput *et al.*, 2007), and to be involved in recycling muropeptides during cell elongation and/or division (Engel *et al.*, 1991, Chaput *et al.*, 2007). Slt has previously been reported as a virulence factor (Mushegian *et al.*, 1996), and the down-regulation of this gene by *F. novicida* in response to elevated temperature was unexpected, although one possible explanation could be that *F. novicida* reinforces its cell membrane in response to elevated temperature, reducing the availability of surplus muropeptides for recycling. This hypothesis may be supported by a recent publication where it was demonstrated that a *H. pylori* Δ*slt* mutant accumulated extra long glycan strands in its peptidoglycan layer, suggesting a role for *slt* in regulation of glycan length (Chaput *et al.*, 2007).

After one hour at the elevated temperature, *galU*, the gene encoding UTP-glucose-1-phosphate uridylyltransferase, was up-regulated by *F. novicida*. In *E. coli* the functional enzyme, a tetramer of the *galU* gene product, converts UTP-glucose to UDP-glucose, which is in itself of central importance in the synthesis of components of the *E. coli* cell envelope (Weissborn *et al.*, 1994). UDP-glucose, in particular that formed through GalU catalysis, is also required for the biosynthesis of LPS, capsular polysaccharide, and membrane-derived oligosaccharides (Dean and Goldberg, 2002; Chang *et al.*, 1996; Nesper *et al.*, 2001; Weissborn *et al.*, 1992) and the up-regulation of this gene may further indicate that *F. novicida* modifies one or more of these cell membrane components in response to elevated temperature.

Chapter 5d

The acidic pH response of *F. novicida*

5d.1 Introduction

5d.1.1. Is the *Francisella*-containing phagosome acidified?

It was reported by Fortier *et al.*, that *F. tularensis* LVS grows in an acidified compartment which provides essential iron that is required for growth because the low pH causes the release of iron from transferrin (Fortier *et al*, 1995). However in direct contrast to these findings, Clemens *et al.* have more recently reported that endosomes containing live *Francisella* are at pH 6.7 and therefore not significantly acidified (Clemens *et al.,* 2004). This group propose that transferrin is delivered to the early phagosome (approximately 15 minutes after infection), but no explanation is offered as to how iron is released from transferrin at neutral pH. Clemens *et al.* also calculated that killed bacteria reside in acidified phagosomes at pH 5.5, indicating that *F. tularensis* possesses some mechanism for preventing acidification of the phagosome (Clemens *et al.*, 2004). In the Clemens study, the phagosomal pH of human monocytes at approximately three hours post-infection with *F. tularensis* subsp. *tularensis* was measured directly using an immunoelectron microscopy technique. In the Fortier study, the acidification of the phagosome was inferred from a reduction in viable *F. tularensis* LVS recovered at 72 hours post-infection from murine monocytes that were treated with a variety of agents selected for their ability to prevent endosome acidification. An obvious discrepancy between these two studies is that one was carried out using murine cells infected with *F. tularensis* LVS, whereas the other was carried out using human cells infected with *F. tularensis* subsp. *tularensis*. Therefore perhaps these experiments indicate a difference in the way that human and murine monocytes react to *Francisella*, or perhaps the

difference between the two findings indicates that an ability to prevent acidification of the initial endocytic compartment is a trait of virulent *F. tularensis* strains only. However, in light of the recent report that *F. tularensis* enters the autophagic pathway after replication in the cytosol (Checroun *et al.*, 2006), the possibility must be considered that the anti-acidification reagents used by Fortier *et al.* actually prevented acidification of the FCV rather than of the phagosome, especially as bacteria were enumerated at 72 hours post-infection, meaning that bacteria would have had time to escape the phagosome, although it was assumed in the report that this did not occur. It therefore seems unlikely that the *Francisella*-containing phagosome is acidified, although it does seem likely that *F. tularensis* encounters acidic conditions at some point inside the macrophage.

5d.1.2. Aim

The aim of this study was to use the *F. tularensis* Schu S4 microarray to examine the transcriptomic profile of *F. novicida* as a specific response to acidic pH. Determination of the response of *F. novicida* to low pH can provide clues as to whether similar conditions are encountered by *Francisella in vivo*, and clues as the survival mechanisms that *Francisella* employs when confronted with such conditions.

5d.2. Results

5d.2.1. Growth of *F. novicida* at reduced pH

It has previously been reported that, at three to four hours post-infection of human macrophages, live *F. tularensis* subsp. *holarctica* reside in a phagosome with a pH of 6.7, whereas the pH of phagosomes containing formalin-killed *F. tularensis* was measured as pH 5.5 (Clemens *et al.*, 2004). The pH of CDM produced for this study was measured as 6.2, and, as this slightly acidic medium had supported the growth *F. novicida* in previous studies, this was used as the control compared to acidified CDM at pH 5.4. In order to select biologically relevant sampling times for transcriptomic studies, a comparison of the growth rate of F. novicida cultured at pH 5.4 and at pH 6.2 (the control) was compared. Adjustment of the pH from 6.2 to 5.4 did not affect the growth rate of *F. novicida* (fig. 5d.1).

5d.2.2. The transcriptomic response of *F. novicida* to acidic conditions

Due to the similarity between the growth rate of *F. novicida* at neutral and acidic pH (fig. 5d.1), the transcriptomic response of *F. novicida* to acidic pH was investigated at three time points selected during the first hour post stress-induction: 15 min., 30 min., and 60 min. As described in section 5.1, *F. novicida* was cultured in CDM-Met for transcriptomics analysis in response to acidic pH, RNA was isolated for each condition on three separate occasions (biological replicates) and hybridised to the microarray twice (technical replicates). A list of genes that were regulated by *F. novicida* in response to acidic conditions can be found in table 5d.1. As previously described in section 5a.3.2.1., as the microarray oligonucleotides used to

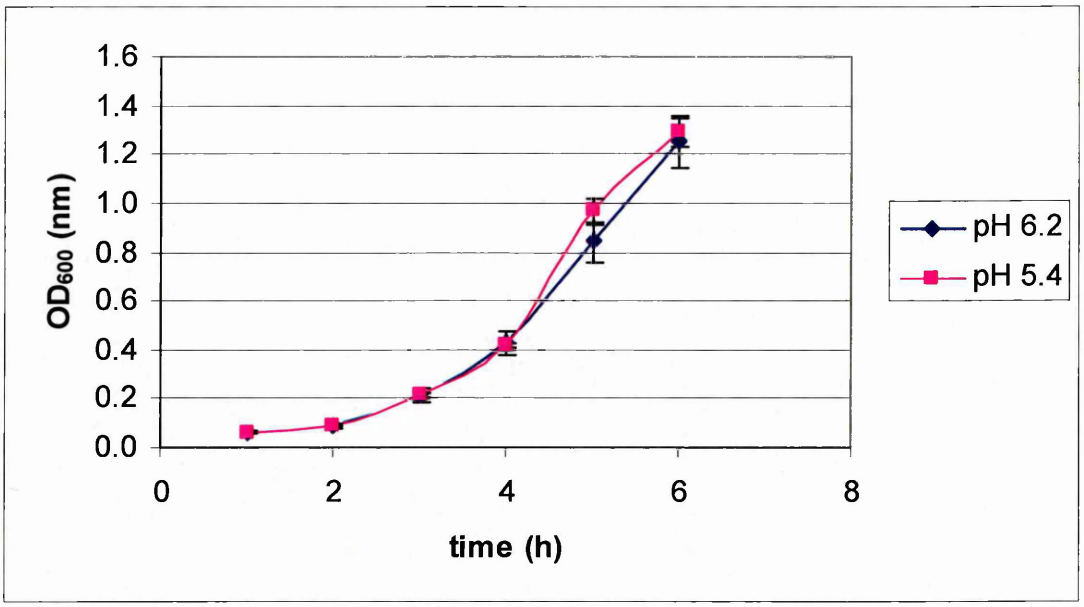


Fig. 5d.1 Comparison of the replication rate of *F. novicida* cultured in CDM at pH 5.4 vs. pH 6.2. The growth rate was equivalent for all of the time points taken.

Table 5d.1 Genes regulated by *F. novicida* in response to culture under acidic conditions. [a]Locus tag assigned in the *F. novicida* U112 genome sequence (Brittnacher *et al.*, 2006). [b]Locus tag assigned in the *F. tularensis* Schu S4 genome sequence (Larsson *et al.*, 2005). [c]Pseudogene in *F. tularensis* Schu S4.

ID	FTN no.[a]	*FT no.*[b]	Time	Fold change	P_{FDR}	Annotation	GenProtEC Classification
Up							
FTN_1068	1068	*FTT0613*	15	1.56	0.05	Hypothetical protein	[13]
fabH	1337	*FTT1373*	15	1.48	0.03	3-oxoacyl-[acyl-carrier protein] synthase III	[1.5.4] [7.1]
FTN_0588	0588	*FTT0497*[c]	30	1.45	0.05	Asparaginase	[1.7.13]
FTN_1345	1345	*FTT1381*[c]	30	1.45	0.05	Hypothetical protein	[13]
Down							
FTN_1024	1024	*FTT0659*	15	2.13	0.04	RmuC family protein	[12]
hflK	1048	*FTT0633*	15	1.46	0.05	HflK-HflC membrane protein complex, Hflk	[8.1] [11]
FTN_0080	0080	*FTT0240*	30	1.45	0.05	SAM-dependent methyltransferase	[12]
FTN_0086	0086	*FTT0225*[c]	30	1.64	0.05	Metabolite H^+ symporter (MHS) family protein	[4.2.A.?]
crcB	0136	*FTT0260*	30	1.46	0.05	CrcB family protein	[5.1] [6.1]
FTN_0300	0300	*FTT1629*	30	1.51	0.05	Glycosyl transferase, group 2	[12]
FTN_0494	0494	*FTT0398*	30	1.73	0.05	Hypothetical membrane protein	[11] [12]
FTN_0904	0904	*FTT1026*	30	1.60	0.05	Hypothetical protein	[13]
rplT	1188	*FTT0820*	30	1.47	0.05	50S ribosomal protein L20	[2.3.2] [2.3.8] [11]
FTN_1210	1210	*FTT0801*	30	2.55	0.05	Ribokinase, PfkB family	[1.1.1] [1.3.1] [7.1]
FTN_1406	1406	*FTT1438*[c]	30	1.50	0.05	Hypothetical protein	[13]
lysA	1530	*FTT0027*	30	1.48	0.05	Diaminopimelate decarboxylase	[1.5.1.7] [11]
FTN_0544	0544	*FTT0453*	60	7.33	0.05	Hypothetical protein	[11] [13]
FTN_1160	1160	*FTT1182*[c]	60	11.63	0.05	VacJ-like lipoprotein	[6.1] [7.3] [11]

carry out *F. novicida in vitro* stress response studies were originally designed to be complementary to the *F. tularensis* Schu S4 genome sequence, the sequence of each oligonucleotide that was identified as regulated by *F. novicida* to elevated temperature was compared to the *F. novicida* genome sequence that was subsequently available (Brittnacher *et al.*, 2006) using the BLASTn program (Altschul *et al.*, 1990). In most cases 100% complementarity was achieved between the oligonucleotide and the *F. novicida* genome sequence and within the *F. novicida* CDS corresponding to the *F. tularensis* Schu S4 CDS to which the oligonucleotide was designed. If less than 100% complementarity between the oligonucleotide and the *F. novicida* genome was achieved, then the most likely *F. novicida* binding site was selected according to the criteria of ≥ 75% overall complementarity and ≥ 15 base pairs contiguous complementarity (Kane *et al.*, 2000). In these cases the sequence of the *F. novicida* CDS selected was then used to find the corresponding CDS in *F. tularensis* Schu S4 using tBLASTx (Altschul *et al.*, 1990) to ensure continuity when interpreting the results in the context of either genome.

5d.2.3. Regulated CDS were grouped according to function

Each of the CDS regulated by *F. novicida* in response to acidic pH was assigned to a functional category as previously defined by the GenProtEC database (Serres *et al.*, 2004) (Appedix 1), and represented graphically (fig. 5d.2). The GenProtEC classification assigns gene products to groups according to their broad function, and into sub-groups which

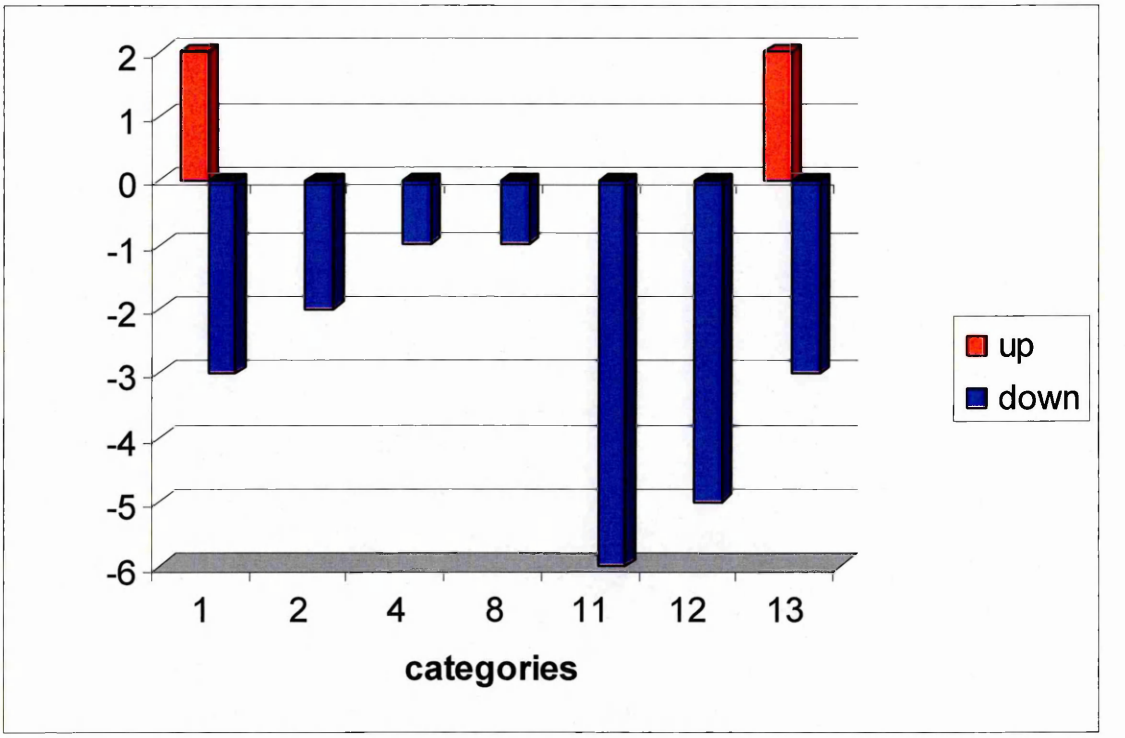


Fig. 5d.2 GenProtEC categories of CDS regulated by *F. novicida* in response to culture under acidic conditions. The x-axis represents GenProtEC categories of CDS regulated in response to acidic conditions were Metabolism (category 1), Information transfer (category 2), Transport (category 4), extrachromosomal genes (category 8), Pathogenic related genes (category 11), Conserved hypothetical genes (category 12), and Hypothetical genes (category 13). The y-axis represents the number of genes up-regulated (+) or down –regulated (-).

describe increasingly specific functions at each level. It should be noted that some CDS are assigned to more than one GenProtEC category.

5d.2.3.1. Functional groups of up-regulated CDS

The four CDS that were up-regulated by *F. novicida* in response to acidic pH were divided equally between the Metabolism and Hypothetical groups (fig. 5d.2). The up-regulated CDS from the Metabolism functional group were linked to amino acid biosynthesis and to amino acid conversion (fig. 5d.3). Only one of the final gene products of the four up-regulated CDS was assigned to a cellular location, which was the cytoplasm. None of the four CDS up-regulated in response to low pH were assigned to cellular processes or to a cell structure.

5d.2.3.2. Functional groups of down-regulated CDS

The functional group containing the largest number of CDS down-regulated by *F. novicida* in response to acidic pH was Pathogenic-related genes, followed by the two Hypothetical groups (fig. 5d.2). Three down-regulated CDS were assigned to the Metabolism group (fig. 5d.3), including carbon compounds metabolism, glycolysis, and lysine biosynthesis. The gene products of two CDS were assigned to cellular locations: the cytoplasm and the inner membrane. Two of the down-regulated CDS were linked to the membrane structure, and one was linked to cell division.

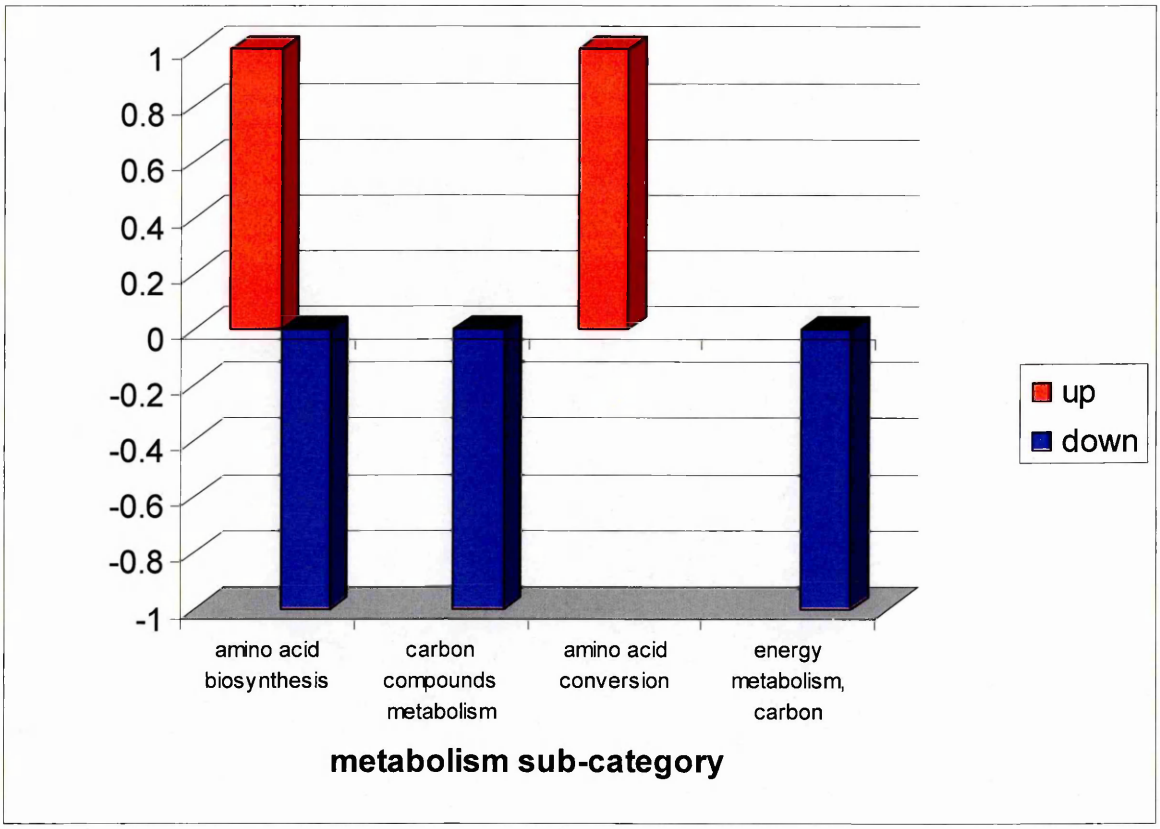


Fig. 5d.3 GenProtEC Metabolism sub-categories of CDS regulated by *F. novicida* in response to culture under acidic conditions (x-axis). The y-axis represents the number of genes up-regulated (+) or down –regulated (-).

5d.2.4. QPCR

TaqMan QPCR was carried out on three CDS that were indicated by microarray as regulated by *F. novicida* in response to acidic pH: FTT0613 was up-regulated, and FTT0659 and *hflK* were down-regulated (table 5d.1). For each RNA sample, the previously reported *F. tularensis* housekeeping gene, *prfB*, was also assayed to provide a normalisation standard (Nübel *et al.*, 2006). It was noted that the level of expression of *prfB* as determined by microarray was similar in both test and control cultures, supporting the use of this as a housekeeping gene. These microarray results were confirmed by QPCR with a Pearson correlation of 0.998 (fig. 5d.4). As mentioned in section 5a.3.4., the overall Pearson correlation between microarray data and TaqMan QPCR data for 18 genes assayed, selected because they were indicated by microarray as regulated by *F. novicida* in response to one of the conditions tested was 0.533.

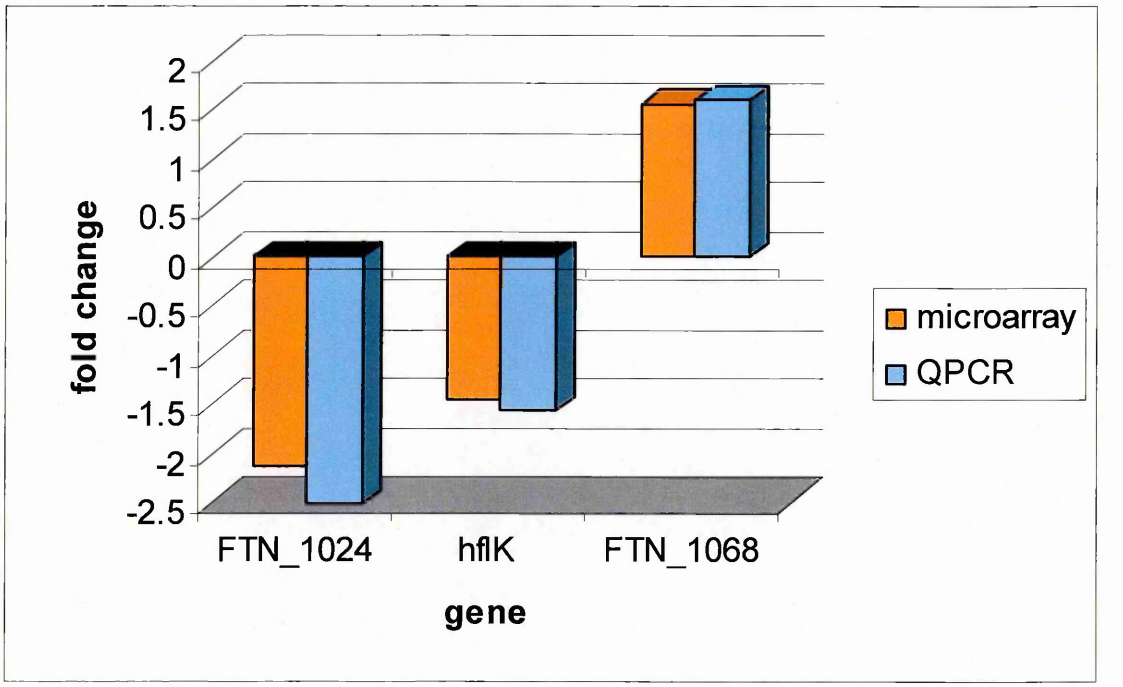


Fig. 5d.4 QPCR confirmation of microarray data. The QPCR data supported the microarray data in that all genes were regulated in the same direction, although not always to the same magnitude:

Gene	Microarray	QPCR
FTN_1024	-2.13	-2.5
hflK	-1.46	-1.56
FTN_1068	1.56	1.6
Correlation	**0.99756586**	

5d.3 Discussion

5d.3.1. The transcriptomic response to a low pH environment

CDS from the GenProtEC Metabolism sub-categories of fatty acid biosynthesis and amino acid conversion were up-regulated by *F. novicida* in response an acidic pH, the implications of up-regulation of both of these CDS are discussed below. A number of CDS assigned to GenProtEC Pathogenic-related genes were down-regulated by *F. novicida* at low pH, possibly suggesting that the intracellular pH is higher than the experimental pH tested here.

5d.3.2. Many of the genes up-regulated at low pH are annotated as hypothetical

One of the two hypothetical proteins up-regulated in response to acidic conditions, FTN_1068, was also up-regulated by *F. novicida* in response to culture at 42°C. This gene has been shown in *F. novicida* to be under the regulatory control of a recently described orphan response regulator, encoded by *pmrA* (Mohapatra *et al.*, 2007). The PmrAB two component regulator is a well characterised bacterial virulence factor (Soncini and Groisman, 1996; Marchal *et al.*, 2004). It has been shown to confer resistance to polymyxin B (Soncini and Groisman, 1996), and the expression of the response-regulator PmrA in particular has been shown to be induced by mildly acidic pH through a mechanism involving protonation of PmrB (Soncini and Groisman, 1996; Perez and Groisman, 2007). It is thought that PmrA regulates a number of genes which in turn modify the lipid A portion of LPS, altering the barrier properties of the outer membrane (Marchal *et al*,

2004; Gibbons *et al.*, 2005; Froelich *et al.*, 2006). A Δ*pmrA* mutant of *F. novicida* was defective for growth in both murine and human macrophage lines, and was attenuated for virulence in a murine model of infection (Mohapatra *et al.*, 2007). The up-regulation of FT_1068 in response to two stress conditions, coupled with the fact that it is regulated by a virulence factor that is known in turn to be activated by mildly acidic pH, may indicate that expression of this gene is important for *Francisella* infection in mammals, and in particular under acidic conditions that may be encountered in the intracellular environment.

5d.3.3. A fatty acid biosynthesis gene was up-regulated at low pH

One gene that is involved in fatty acid biosynthesis, *fabH*, was also up-regulated by *F. novicida* in response to acidification of the growth medium. This gene encodes a 3-oxoacyl-[acyl-carrier protein] synthase III, which may indicate that *F. novicida* fortifies itself against the lower pH by altering or increasing the fatty acid component of its outer membrane. None of the other fatty acid biosynthesis pathway genes were up-regulated in response to the lower pH (fig. 5d.5), but if an increase in the amount of fatty acid available is required, it is possible that the gene product of *fabH* is the limiting component under non-stressful conditions. Interestingly, *fabH* has been shown to be up-regulated by the acidophilic gastric bacterium *H. pylori* upon attachment to human stomach epithelial cells (Kim *et al.*, 2004), which further supports the role of this gene in survival of bacteria subject to low pH and perhaps also suggests a role in infection if it is true that acidic conditions are encountered inside the macrophage (section 5d.1.1.).

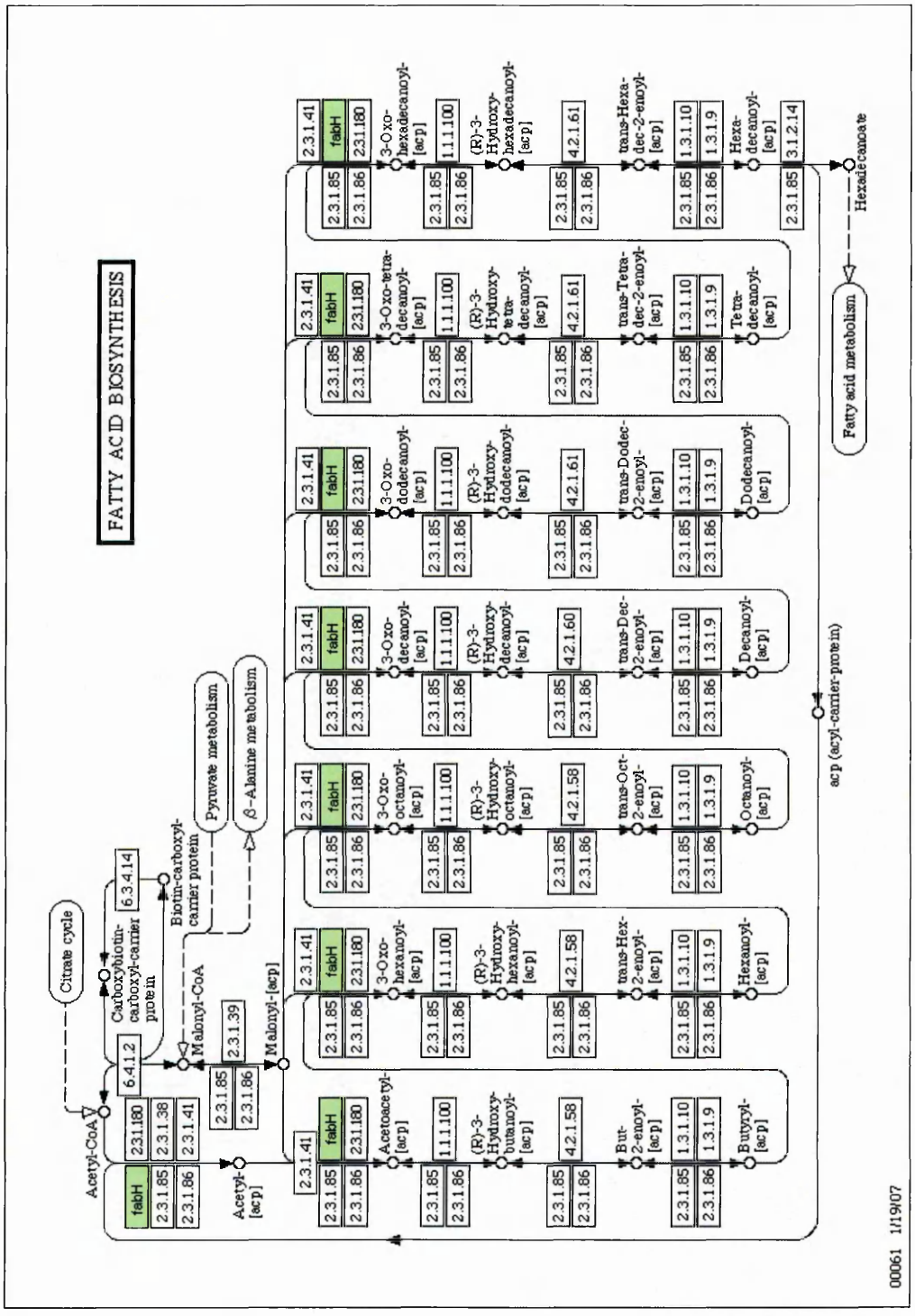


Fig. 5d.5 Fatty acid biosynthesis pathway (KEGG, 2007). The gene highlighted in green (*fabH*) was up-regulated by *F. novicida* in response to oxidative stress.

5d.3.4. An asparaginase gene was up-regulated at low pH

The gene encoded by FTN_0588, which was up-regulated by *F. novicida* in response to acidic conditions, is annotated as an asparaginase, an enzyme which catalyzes the conversion of asparagine to aspartic acid and an ammonium ion. The up-regulation of an asparaginase by *F. novicida* in response to low pH is in contrast to the acid response of *H. pylori*, which has been shown to down-regulate asparaginase at low pH (Merrell *et al.*, 2003). However, the production of ammonium ions (as a product of urease catalyzed hydrolysis of urea) has been shown to protect both *H. pylori* and *Y. pseudotuberculosis* under acidic conditions. Ammonium ions raise the local pH, acting to buffer the acidic environment (Marshall *et al.*, 1990; De Koning-Ward & Robins-Browne, 1995), and therefore it is intuitive that up-regulation of an enzyme which catalyses a reaction with ammonium as a product is likely to be a mechanism by which *F. novicida* protects itself when faced with acidic conditions.

5d.3.5. A number of virulence-associated genes were down-regulated at low pH

A VacJ-like lipoprotein, encoded by FTN_1160, was down-regulated by *F. novicida* at the lower pH. VacJ is involved in cell envelope biosynthesis and has previously been showed to be up-regulated by *Pseudomonas putida* in response to stress by phenol (Santos *et al.*, 2004). VacJ is also associated with the intercellular spread of *S. flexneri* (Suzuki *et al.*, 1994) and has been shown by subtractive hybridisation to be present in the highly pathogenic *N. meningitidis* but absent in less pathogenic *N. gonorrhoeae*

(Perrin *et al.*, 1999). It may be pertinent to note that *vacJ* is a pseudogene in *F. tularensis* strain Schu S4 and therefore cannot contribute to the high virulence of this strain. However, *F. novicida* is virulent in some species, for example mice, and the possibility cannot be discounted that this gene is a virulence factor for this species of *Francisella*. Also down-regulated in response to low pH were two genes that have been shown to play a role in *F. tularensis* LVS infection of the mouse lung via the intranasal route: *hflK* and *rplT* (Su *et al.*, 2007), and three genes that have been shown to be required for the survival of *F. novicida* in murine spleens after i.p. infection: FTN_0494, *lysA*, and FTN_0544 (Weiss *et al.*, 2007). One possible physiological explanation that can be proposed for the down-regulation of *hflK* under stressful conditions is that *hflK* is part of the FtsH enzyme which is an ATP-dependent protease that is universally conserved in bacteria (Bieniossek *et al.*, 2006). In *E. coli* the heat shock response depends on a transient increase in the intracellular level of σ^{32}, FtsH being a major mediator of σ^{32} degradation (Kanemori *et al.*, 1999, and Arsène *et al.*, 2000). It is possible that a similar mechanism exists in *F. tularensis* and that there is a requirement for stabilization of σ^{32} (through down-regulation of *hflK)* in response to acidic conditions.

5d.3.6. These data suggest that the intracellular pH is at times higher than 5.5

The paucity of genes up-regulated specifically in response to low pH conditions, combined with down-regulation of a number of genes that might be expected to play a role in *Francisella* virulence and comparatively similar

growth rates, at least for the first hour, under low pH and control conditions suggests that *F. novicida* is not stressed at the lower pH, and that in fact it is quite able to survive within the pH range investigated here without eliciting a classical stress response. It is important to note that, because these microarray data were used to calculate the *relative* expression between two states, there are two possible interpretations of the data i.e. genes which are down-regulated under test conditions (compared to control) can also be considered to be up-regulated under control conditions (compared to test). Therefore, if the findings of Clemens *et al.* are correct, and phagosomes containing live *F. tularensis* are at pH 6.7 (Clemens *et al.*, 2004), then the experimental control pH of 6.2 is actually closer to the real intraphagosomal situation and these data could actually indicate that putative virulence factors such as *vacJ,* and genes with a demonstrated role in *Francisella* infection, are up-regulated at the pH found *in vivo* compared to the lower pH tested here. Of course this does not preclude the possibility that a lower pH could be encountered by *Francisella* at some other stage of its intracellular journey, including in the autophagosome as suggested by Checroun *et al.* (Checroun *et al.*, 2006).

5.2. Conclusions

The main advantage of culturing *F. novicida* under conditions of methionine starvation, so that the growth phases of both the test and control samples would remain the same, may also have proven to been the main disadvantage, in that a reduction in growth rate, probably through induction of the SOS response, in fact forms an integral part of the bacterial response to stressful conditions. Methionine was selected for removal from the growth media because *F. tularensis* LVS has previously been shown to be able to grow for at least one hour in the absence of this amino acid, including whilst under either heat shock or oxidative stress conditions, in order for investigators to pulse-label samples using [^{35}S]methionine (Ericsson *et al.*, 1994; Golovliov *et al.*, 1997). The main risk of using this strategy was that genes which may have been regulated in response to the main condition of interest were also regulated by bacteria for which both the test and control cultures were starved of methionine, and therefore were not observed by a technique such as microarray in which transcriptomes are compared. In fact this would appear to have been the case, as indicated by the low number of changing genes identified under the three conditions for which methionine-free media was used (elevated temperature, low pH, and oxidative stress [22, 18, and 21 genes, respectively]), compared to 53 changing genes identified under iron-free growth, which was carried out using methionine-replete media. However, the benefit of using this approach was that genes identified as regulated by *F. novicida* in response to elevated temperature, to low pH and to oxidative stress were highly specific to the conditions of interest and insomuch were not incorrect, and in fact the rate of false-positives was so far

reduced as to probably provide a number of false negatives. This could be viewed as a good thing from a technique which can traditionally be thought of as generating unwieldy amounts of data with a high rate of false positives. However, culture under starvation conditions would not be recommended for future studies.

In this study a number of genes were regulated by *F. novicida* in response to *in vitro* stress conditions that have been shown previously to be required for *Francisella* infection (table 5.1). Some genes, or genes with similar suggested functions, were up-regulated in response to one condition but down-regulated in response to another, for example *lysA* and *frgA*, which form part of the *F. tularensis* siderophore operon, were down-regulated in response to hydrogen peroxide but up-regulated in response to iron-starvation. It is possible to provide a possible explanation for both of these results individually, but, as both iron-starvation and ROS are likely to be encountered *in vivo*, these types of conflicting results can be difficult to interpret with respect to a real infection. This is a real limitation of *in vitro* expression studies, and demonstrates the value of data obtained using *in vivo* transcription models.

5.2.1. Future work

The nature of this study has been to investigate the transcriptome of *F. novicida* under selected conditions that are usually thought of a stressful to intracellular bacteria. Information about genes that are regulated by bacteria during infection can provide targets for vaccines, either through the

Table 5.1 Genes regulated by *F. novicida* in response to stress that have previously shown to be attenuating in *Francisella* spp.

Regulation by *F. novicida*	Gene ID	FTN no.	FTT no.	Function	Impact of mutation on virulence
Up-regulated under iron-starvation	*frgA*	1682	0029	Siderophore	Attenuating in *F. tularensis* LVS: mouse model of infection (i.p. route) (Su *et al.*, 2007)
	lysA	1530	0027		Attenuating in *F. novicida*: mouse model of infection (i.p. route) (Weiss *et al.*, 2007)
	FTN_1683	1683	0028		
	FTN_1685	1685	0026		
	pyrD	0036	1647	Pyrimidine biosynthesis	
	pyrF	0035	1648		
Down-regulated under iron-starvation	FTN_1256	1256	1238	Membrane protein	
	hisS	1658	0052	Histidyl-tRNA synthetase	
	kdpD	1715	1736	Regulator	
	wbtE	1426	1460	LPS O-antigen biosynthesis	
	wbtH	1421	1456		
	clpB	1743	1769	Chaperone	Attenuating in *F. tularensis* LVS and *F. novicida*: mouse models of infection (i.p. route) (Su *et al.*, 2007; Weiss *et al.*, 2007)
	pdpD	1325	1360	Hypothetical: FPI encoded	Attenuating in *F. novicida*: mouse *in vivo* and *ex vivo* models of infection (Nano *et al.*, 2004)
Up-regulated at 42°C (vs. 37°C)	FTN_1326	1326	1361	Hypothetical	Attenuating in *F. novicida*: mouse model of infection (i.p. route) (Weiss *et al.*, 2007)
	sdhC	1639	0072	Oxidative phosphorylation	
Down-regulated at pH 5.4 (vs. pH 6.2)	FTN_0494	0494	0398	Hypothetical membrane	
	lysA	1530	0027	Siderophore	
	FTN_0544	0544	0453	Hypothetical	
	hflK	1048	0633	Membrane protein	Attenuating in *F. tularensis* LVS: mouse model of infection (i.p. route) (Su *et al.*, 2007)
	rplT	1188	0820	50S ribosomal protein	

generation of rationally attenuated strains or through the provision of potential antigens for subunit vaccines. At the time of writing, seven genes regulated by *F. novicida* in this study have been forwarded to a Tularemia Vaccine programme for further investigation as potential vaccine candidates.

It has been possible to suggest reasons as to why *F. novicida* regulates some of the observed genes in response to these conditions based on either the putative or confirmed functions of the genes and on previously published works. However, in order to confirm or disprove these hypotheses, and perhaps to put them into the context of a fully virulent *F. tularensis* strain, further experimental work is essential, for example the generation and characterisation of deletion mutants in genes shown to be regulated in response to a given condition mentioned previously. Characterisation could include investigation of the response of the mutant strain to the condition under which gene regulation was observed here, or to other conditions designed to mimic an infection environment. Responses other than transcription that could also be measured, for example proteomics techniques could be used to study gene expression, or specific assays could be designed to measure the output of a particular gene, both in mutant and wild type strains. Use of model systems of infection, both *in vivo* and *ex vivo*, would also be excellent ways to characterise mutant strains.

Of particular interest from the iron-starvation response data presented here was the second putative iron-related cluster of genes that were up-regulated by *F. novicida* in response to iron-starvation (section 5a.4.3.), deletion mutants could be constructed in any or all of these genes and characterised

using some of the techniques mentioned above, and in particular in response to iron-starvation. The response of *F. novicida* to H_2O_2 cannot be fully characterised without transcription data for *katG* and *sodB*, as roles for these two genes in the oxidative stress response of *F. tularensis* have been previously reported (section 5b.4.1.). Therefore work to further characterise the oxidative stress response of *F. novicida* should investigate these two genes in particular. Based on the data presented here, it would also be interesting to perform further experiments to assess whether pyruvate is important in the response of *Francisella* to oxidative stress brought about by H_2O_2 (section 5b.4.2). It would also be interesting to see whether a *Francisella* ΔFTN_1068 strain was able to survive under stress conditions and in particular *in vivo*. This ORF was up-regulated under both elevated temperature and under acidic pH conditions and has previously been demonstrated to be under the control of the known virulence regulator PmrA (section 5d.4.2) meaning that this could be an important *Francisella* stress response gene.

Finally, a most important study that could be carried out using the *F. tularensis* DNA microarray would be to investigate the transcription profile of *Francisella in vivo*. However, this type of experiment is technically challenging in a variety of ways, not least because an *in vivo* (or *ex vivo*) model contains a large amount of contaminating host mRNA compared to a relatively low amount of bacterial mRNA. There are two main approaches often taken to solve this problem: separation of the bacterial cells from the host cells in advance of bacterial cell lysis, thus preventing mixing of the two

mRNA populations, or concurrent lysis of both host and bacterial cells with enrichment for bacterial mRNA. The first approach, separation of the two cell populations, requires an agent that will selectively lyse host cells whilst leaving the bacteria intact. This has been successfully achieved for other bacterial species using *ex vivo* models, for example *M. tuberculosis*-infected macrophages can be lysed using guanidium isothiocyanate, but the complex outer membrane of this acid-fast bacterium protects it from lysis (Butcher *et al.*, 1998). This lysis agent also serves to stabilize the RNA from further transcription and degradation (Butcher *et al.*, 1998). However, bacteria that do not have a complex outer membrane, including *Francisella*, will not be protected from lysis by guanidium isothiocyanate. The approach taken to solve this problem for *S. enterica* serovar Typhimurium was to use a mixture of SDS, phenol, and ethanol (Eriksson *et al.*, 2003). SDS was the lysis agent whilst the phenol and ethanol mixture served to stabilize the RNA (Eriksson *et al.*, 2003). A mixture of saponin and RNAprotect (Ambion) has also been recommended for the same purpose (Ambion, 2003). However, whilst an *ex vivo* model can provide valuable insights into the transcription profile of bacteria upon infection, it cannot fully substitute for an *in vivo* model of infection. Selective lysis of host cells as tissue, i.e. from an *in vivo* model, is more problematic than for tissue cultured cells because bacteria are not able to withstand any agent that is able to lyse whole or even ground tissues. For transcription profiling of bacteria from *in vivo* models of infection it is more often reported that both host and bacterial cells are lysed, with separation of the two RNA populations post-lysis. Separation is sometimes carried out using poly-thymine-based affinity chromatography, on the basis that

prokaryotic mRNA is not polyadenylated, for example using the MICROBEnrich system (Ambion). However, the assertion that only eukaryotic mRNA is polyadenylated is not correct, and it has been reported that at any given time up to 50% of bacterial mRNA may be polyadenylated (Sarkar, 1997). Another approach used with mixed populations of RNA is to enrich for bacterial mRNA using specific primers for RT-PCR (Talaat *et al.*, 2000). A potential limitation of these genome-directed primers is that, unlike random primers which inevitably bind at multiple sites within each ORF, genome-direct primer sets are designed to include only the minimum number of oligonucleotides of a specified length required to bind to each ORF once. This means that failure of any primer to bind to one or more of its corresponding ORFs will bias the data set against these genes. Both *in* and *ex vivo* models of *Francisella* infection have been developed at Dstl, and work to refine *Francisella* transcription profiling techniques using these models is underway. Transcription profiling using the *ex vivo* model has focussed on separation of the two intact cell populations, with lysis using deoxycholate and concurrent RNA stabilisation using RNAprotect. The main limitation of this model has been low numbers of bacteria recovered, even at the optimal point of infection for maximum recovery (data not shown). Transcription profiling using the *in vivo* model has focussed on enrichment of bacterial mRNA from a mixed population using genome-directed primers. However, this enriched bacterial mRNA may still be too contaminated with host RNA for use with a DNA microarray (data not shown). Future work to solve this problem may include further refining a *Francisella* genome-directed primer set to specifically exclude priming to host RNA (Blick *et al.*, 2003), or

enrichment for bacterial mRNA based on the polyadenylation status of host mRNA, accepting the limitations of this technique.

Appendix 1

RD observed by aCGH in *Francisella* strains

Appendix 1 RD observed in *F. tularensis* strains

Appendix 1.1 RD observed in *F. tularensis* subsp. *holarctica*, FSC012

Gene ID	Gene ID	Gene ID	Gene ID	Gene ID	Gene ID	Gene ID
16s rRNA	FTT0158	FTT0768	FTT1159	FTT1693	*lipA*	*rplV*
acnA	FTT0170	FTT0784	FTT1162	FTT1761	*lpxA*	*rpmA*
acpA	FTT0200	FTT0823	FTT1180	FTT1763	*lpxC*	*rpoA1*
alaS	FTT0210	FTT0825	FTT1184	FTT1776	*maeA*	*rpsF*
aroG	FTT0219	FTT0845	FTT1185	FTT1779	*malP*	*rpsG*
atpH	FTT0230	FTT0849	FTT1198	FTT1783	*mdaB*	*rpsU3*
bioD	FTT0242	FTT0851	FTT1201	FTT1791	*mdh*	*sdhC*
bioF	FTT0244	FTT0865	FTT1203	FTT1799	*merA*	*serA*
birA	FTT0248	FTT0898	FTT1211	FTT1804	*metK*	*sun*
clpX	FTT0254	FTT0902	FTT1242	*ftsA*	*moxR*	*talA*
cynT	FTT0272	FTT0909	FTT1252	*ftsQ*	*msrA1*	*thyA*
cyoC	FTT0292	FTT0913	FTT1277	*gidA*	*mutM*	*tmk*
cysC	FTT0294	FTT0921	FTT1292	*glgA*	*napH*	*trkA*
dedA2	FTT0301	FTT0931	FTT1308	*glgC*	*nupC*	*trmD*
def2	FTT0310	FTT0933	FTT1332	*glk1*	*nusG*	*trpG*
dgt	FTT0376	FTT0949	FTT1334	*glmU*	*omp26*	*trpG1*
dinP	FTT0394	FTT0961	FTT1343	*glnA*	*oppC*	*trpR*
dnaA	FTT0401	FTT0965	FTT1350	*gloA*	*oppF*	*trpS*
dnaB	FTT0433	FTT0973	FTT1361	*glpK*	*panC*	*trxA1*
dnaG	FTT0444	FTT0975	FTT1380	*glyQ*	*panD*	*tsf*
dnaN	FTT0446	FTT0981	FTT1388	*gpmI*	*parB*	*ung*
engA	FTT0487	FTT0991	FTT1399	*groL*	*pdpD*	*usp*
eno	FTT0515	FTT0996	FTT1423	*grxB*	*pdxY*	*valB*
fabG	FTT0519	FTT1012	FTT1426	*gshA*	*pepN*	*valS*
fabI	FTT0531	FTT1023	FTT1439	*gyrA*	*pgsA*	*vanY*
fabZ	FTT0539	FTT1025	FTT1441	*hemE*	*pheS*	*wbtD*
fdh	FTT0550	FTT1043	FTT1492	*hemH*	*phnA*	*wbtG*
feoA	FTT0576	FTT1048	FTT1493	*hflX*	*phoH*	*wbtI*
feoB	FTT0599	FTT1051	FTT1495	*hfq*	*phrB*	*wbtJ*
ffh	FTT0602	FTT1071	FTT1497	*hipA*	*plsX*	*xthA*
folK	FTT0604	FTT1072	FTT1522	*htpG*	*polA*	*yagD*
frr	FTT0611	FTT1073	FTT1536	*iciA*	*ppiC*	*yajR*
FTT0003c	FTT0628	FTT1075	FTT1565	*iglA*	*ppnK*	*yhbH*
FTT0012	FTT0663	FTT1077	FTT1582	*ilvD*	*ppx*	*yjjV*
FTT0024	FTT0667	FTT1078	FTT1586	*infC*	*prfC*	
FTT0045	FTT0669	FTT1079	FTT1588	*ISFtu1*	*priA*	
FTT0046	FTT0696	FTT1080	FTT1589	*ISFtu2*	*proS*	
FTT0095	FTT0718	FTT1082	FTT1594	*lspA*	*purH*	
FTT0096	FTT0719	FTT1083	FTT1598	*ispF*	*pyk*	
FTT0097	FTT0732	FTT1089	FTT1650	*kdpA*	*pyrB*	
FTT0101	FTT0737	FTT1109	FTT1651	*kdpC*	*recX*	
FTT0105	FTT0744	FTT1136	FTT1659	*kdsA*	*rne*	
FTT0112	FTT0747	FTT1140	FTT1667	*kdsB*	*rplK*	
FTT0156	FTT0748	FTT1141	FTT1684	*kdsB*	*rplO*	

Appendix 1.2 RD observed in *F. tularensis* subsp. *holarctica*, FSC338.

Gene ID	Gene ID	Gene ID
16s rRNA	FTT0965	*glyQ*
acnA	FTT1068	*grxB*
acpA	FTT1069	*gyrA*
alaS	FTT1070	*ilvD*
aroG	FTT1071	*lipA*
cynT	FTT1072	*napH*
dinP	FTT1073	*nusG*
ffh	FTT1242	*oppF*
folK	FTT1252	*panC*
FTT0095	FTT1277	*pheS*
FTT0112	FTT1308	*ppx*
FTT0177	FTT1334	*purH*
FTT0446	FTT1425	*recX*
FTT0488	FTT1426	*res*
FTT0552	FTT1443	*sdhC*
FTT0747	FTT1580	*usp*
FTT0851	FTT1791	*wbtJ*
FTT0961	*glgC*	*xthA*

Appendix 1.3 RD observed in *F. tularensis* subsp. *holarctica*, FSC155.

Gene ID	Gene ID	Gene ID
16s rRNA	FTT1068	*hfq*
acnA	FTT1069	*ilvD*
alaS	FTT1070	*kdpC*
aroG	FTT1071	*lipA*
cynT	FTT1072	*napH*
dedA2	FTT1073	*nusG*
dinP	FTT1242	*oppF*
ffh	FT1252	*ostA1*
folK	FTT1308	*panC*
FTT0095	FTT1334	*pheS*
FTT0112	FTT1425	*ppx*
FTT0156	FTT1426	*prfC*
FTT0177	FTT1443	*purH*
FTT0446	FTT1580	*recX*
FTT0488	FTT1791	*res*
FTT0552	*ftsA*	*riml*
FTT0747	*glgC*	*sdhC*
FTT0851	*glyQ*	*usp*
FTT0961	*grxB*	*wbtJ*
FTT0965	*gyrA*	*xthA*

Appendix 1.4 RD observed in *F. tularensis* subsp. *holarctica*, FSC200.

Gene ID	Gene ID
FTT0522	*napH*
FTT0552	*oppF*
FTT0844	*pdpD*
FTT1071	*res*
FTT1308	*thrA*
FTT1361	*xthA*
FTT1791	*yhiP*

Appendix 1.5 RD observed in *F. tularensis* subsp. *holarctica*, FSC354.

Gene ID
FTT0177
FTT0552
FTT0604
FTT0843
FTT0844
FTT0961
FTT1070
FTT1071
FTT1073
FTT1308
FTT1580
hflB
mdaB
oppF
ostA1
pdpD
xthA

Appendix 1.6 RD observed in *F. tularensis* subsp. *holarctica*, FSC352.

Gene ID	Gene ID	Gene ID	Gene ID	Gene ID	Gene ID
apaH	FTT0455	FTT0977	FTT1407	*hipA*	*ribF*
birA	FTT0498	FTT0978	FTT1416	*holB*	*rimK*
blaA	FTT0505	FTT0988	FTT1419	*hsdM*	*rne*
capB	FTT0522	FTT1007	FTT1441	*hslR*	*rplF*
capC	FTT0523	FTT1010	FTT1492	*htpG*	*rplL*
csp	FTT0527	FTT1040	FTT1502	*htpX*	*rplV*
cspC	FTT0539	FTT1045	FTT1511	*htrB*	*rpmB*
cydA	FTT0550	FTT1051	FTT1513	*infB*	*rpmG*
cyoD	FTT0554	FTT1052	FTT1514	*lpnB*	*rpmH*
dacB1	FTT0596	FTT1064	FTT1519	*lpxH*	*rpoB*
dnaA	FTT0599	FTT1066	FTT1525	*ISFtu1*	*rpoC*
dsbB	FTT0603	FTT1069	FTT1557	*ISFtu2*	*rpsF*
fadB/acbP	FTT0611	FTT1071	FTT1566	*ispA*	*rpsQ*
folE	FTT0622	FTT1072	FTT1576	*ispF*	*secB1*
folK	FTT0628	FTT1083	FTT1583	*kbl*	*secB2*
FTT0003c	FTT0645	FTT1102	FTT1624	*kdsB*	*secD*
FTT0024	FTT0647	FTT1113	FTT1633	*lepB*	*secE*
FTT0028	FTT0660	FTT1135	FTT1640	*lolA*	*smpB*
FTT0069	FTT0665	FTT1159	FTT1645	*lolB*	*sohB*
FTT0083	FTT0685	FTT1166	FTT1655	*maeA*	*speH*
FTT0101	FTT0745	FTT1167	FTT1659	*manC*	*sun*
FTT0135	FTT0755	FTT1184	FTT1686	*minE*	*tdk*
FTT0155	FTT0761	FTT1195	FTT1690	*mltA*	*thrA*
FTT0161	FTT0767	FTT1198	FTT1691	*mpl*	*tolC*
FTT0165	FTT0768	FTT1207	FTT1723	*napH*	*tpiA*
FTT0166	FTT0786	FTT1209	FTT1729	*nuoJ*	*treA*
FTT0177	FTT0795	FTT1219	FTT1747	*nusB*	*trpS*
FTT0180	FTT0828	FTT1272	FTT1757	*olmA*	*wbtH*
FTT0200	FTT0843	FTT1282	FTT1770	*oppF*	*wbtI*
FTT0201	FTT0844	FTT1284	FTT1791	*pdpD*	*XerD?*
FTT0237	FTT0853	FTT1308	*fumC*	*pdxY*	*xthA*
FTT0241	FTT0864	FTT1320	*gatB*	*phnA*	*ybhR*
FTT0248	FTT0878	FTT1333	*gcp*	*potH*	*yfhQ*
FTT0264	FTT0919	FTT1334	*gcvH*	*prfC*	*yhiP*
FTT0268	FTT0930	FTT1343	*ggt*	*proS*	*yidC*
FTT0289	FTT0931	FTT1347	*gidB*	*purH*	*yjjK*
FTT0308	FTT0933	FTT1348	*glnA*	*pyrB*	
FTT0359	FTT0944	FTT1350	*guaA*	*rep*	
FTT0361	FTT0949	FTT1355	*hemN*	*res*	

Appendix 1.7 RD observed in *F. tularensis* subsp. *holarctica*, FSC358.

Gene ID	Gene ID	Gene ID	Gene ID
acnA	FTT0596	FTT1234	*ISFtu1*
alaS	FTT0604	FTT1240	*lspA*
aroG	FTT0611	FTT1242	*kdpA*
aspC2	FTT0612	FTT1272	*lgt*
atpF	FTT0619	FTT1277	*lipA*
atpH	FTT0628	FTT1308	*maeA*
bioF	FTT0665	FTT1334	*manC*
clpP	FTT0677	FTT1342	*mdaB*
cutC	FTT0732	FTT1361	*moxR*
cyoC	FTT0748	FTT1384	*napH*
cysC	FTT0843	FTT1388	*nrdA*
dedA2	FTT0844	FTT1402	*nupC*
dgt	FTT0845	FTT1422	*oppF*
dnaA	FTT0890	FTT1426	*ostA1*
dnaB	FTT0898	FTT1492	*pdpD*
dnaG	FTT0921	FTT1504	*pgsA*
dnaX	FTT0928	FTT1546	*phoH*
fadA	FTT0961	FTT1565	*priA*
folK	FTT0974	FTT1576	*recX*
FTT0012	FTT0978	FTT1580	*res*
FTT0024	FTT0982	FTT1651	*riml*
FTT0057	FTT0985	FTT1655	*rne*
FTT0177	FTT1001	FTT1658	*rnhB*
FTT0200	FTT1012	FTT1659	*rplO*
FTT0237	FTT1030	FTT1694	*rplY*
FTT0242	FTT1043	FTT1763	*rpoA1*
FTT0244	FTT1051	FTT1764	*serA*
FTT0248	FTT1055	FTT1770	*talA*
FTT0276	FTT1066	FTT1791	*thrA*
FTT0294	FTT1068	*galP1*	*tmk*
FTT0359	FTT1069	*gidA*	*trpG*
FTT0443	FTT1070	*glpQ*	*trpR*
FTT0446	FTT1071	*greA*	*ubiA*
FTT0507	FTT1072	*hemA*	*wbtD*
FTT0522	FTT1073	*hemD*	*wbtH*
FTT0524	FTT1126	*hemN*	*wbtJ*
FTT0550	FTT1203	*holB*	*xthA*
FTT0552	FTT1207	*iciA*	*yajR*
FTT0558	FTT1219	*ilvE*	*ybhR*
			yqaB

Appendix 1.8 RD observed in *F. tularensis* subsp. *holarctica*, FSC124.

Gene ID	Gene ID	Gene ID	Gene ID
16s rRNA	FTT0737	FTT1242	*hflX*
acpA	FTT0794	FTT1252	*hfq*
atpH	FTT0851	FTT1277	*kdpC*
cysC	FTT0898	FTT1308	*maeA*
dinP	FTT0902	FTT1334	*moxR*
dnaB	FTT0961	FTT1400	*oppF*
dnaG	FTT0982	FTT1426	*ostA1*
frr	FTT0996	FTT1443	*panC*
FTT0156	FTT1043	FTT1536	*pilC*
FTT0230	FTT1066	FTT1580	*ppx*
FTT0241	FTT1068	FTT1582	*prfC*
FTT0242	FTT1069	FTT1583	*res*
FTT0244	FTT1070	FTT1684	*rne*
FTT0433	FTT1071	FTT1791	*sdhC*
FTT0446	FTT1072	*ftsQ*	*tmk*
FTT0552	FTT1073	*gidA*	*trpG1*
FTT0576	FTT1174	*glyQ*	*xthA*
FTT0611	FTT1203	*gpmI*	*yhiP*
FTT0719	FTT1209	*grxB*	

Appendix 1.9 RD observed in *F. tularensis* subsp. *holarctica*, FSC257.

Gene ID	Gene ID
bioF	FTT1308
fabZ	FTT1361
FTT0446	FTT1580
FTT0552	FTT1659
FTT0576	FTT1791
FTT0961	*ftsA*
FTT0982	*mdaB*
FTT1066	*napH*
FTT1068	*panD*
FTT1069	*pdpD*
FTT1070	*purH*
FTT1071	*purU*
FTT1073	*sdhC*
FTT1242	*xthA*
FTT1302	

Appendix 2

GenPRotEC categories

1 metabolism
1.1 carbon utilization
1.1.1 carbon compounds
1.1.2 fatty acids
1.1.3 amino acids
1.1.4 amines
1.1.5 other compounds
1.2 macromolecule degradation
1.2.1 RNA
1.2.2 DNA
1.2.3 proteins/peptides/glycopeptides
1.2.4 polysaccharides
1.3 energy metabolism, carbon
1.3.1 glycolysis
1.3.2 pentose phosphate shunt, oxidative branch
1.3.3 pyruvate dehydrogenase
1.3.4 tricarboxylic acid cycle
1.3.5 fermentation
1.3.6 aerobic respiration
1.3.7 anaerobic respiration
1.3.8 ATP proton motive force interconversion
1.3.9 Entner-Doudoroff pathway
1.4 energy production/transport
1.4.1 electron donor
1.4.2 electron acceptor
1.4.3 electron carrier
1.5 building block biosynthesis
1.5.1 amino acids
1.5.1.1 glutamate
1.5.1.10 glycine
1.5.1.11 serine
1.5.1.12 cysteine
1.5.1.13 phenylalanine
1.5.1.14 tyrosine
1.5.1.15 tryptophan
1.5.1.16 histidine
1.5.1.17 alanine
1.5.1.18 isoleucine/valine
1.5.1.19 leucine
1.5.1.2 glutamine
1.5.1.20 chorismate
1.5.1.21 homoserine
1.5.1.3 arginine

1.5.1.4 proline
1.5.1.5 aspartate
1.5.1.6 asparagine
1.5.1.7 lysine, diaminopimelate
1.5.1.8 threonine
1.5.1.9 methionine
1.5.2 nucleotide
1.5.2.1 purine biosynthesis
1.5.2.2 pyrimidine biosynthesis
1.5.2.3 purine ribonucleotide biosynthesis
1.5.2.4 pyrimidine ribonucleotide biosynthesis
1.5.3 cofactor, small molecule carrier
1.5.3.1 biotin
1.5.3.10 glutathione
1.5.3.11 menaquinone, ubiquinone
1.5.3.12 heme, porphyrine
1.5.3.13 cobalamin (Vitamin B12)
1.5.3.14 enterochelin (enterobactin)
1.5.3.2 folic acid
1.5.3.3 lipoate
1.5.3.4 molybdenum
1.5.3.5 Coenzyme A
1.5.3.6 pyridoxine (vitamin B6)
1.5.3.7 nicotinamide adenine dinucleotide
1.5.3.8 thiamin
1.5.3.9 riboflavin (Vitamin B2), FAD, FMN
1.5.4 fatty acid and phosphatidic acid
1.6 macromolecules (cellular constituent) biosynthesis
1.6.1 phospholipid
1.6.10 lipoprotein
1.6.11 glycoprotein
1.6.12 flagella
1.6.13 fimbria, pili
1.6.15 large molecule carriers
1.6.15.1 cytochromes
1.6.15.2 thioredoxin, glutaredoxin
1.6.15.3 biotin carboxyl carrier protein
1.6.15.4 acyl carrier protein
1.6.2 colanic acid (M antigen)
1.6.3 lipopolysaccharide
1.6.3.1 O antigen
1.6.3.2 core region
1.6.3.3 lipid A

1.6.4 enterobacterial common antigen (surface glycolipid)
1.6.5 K antigen
1.6.6 osmoregulated periplasmic glucan
1.6.7 peptidoglycan (murein)
1.6.9 cytoplasmic polysaccharides
1.7 central intermediary metabolism
1.7.1 unassigned reversible reactions
1.7.10 sugar nucleotide biosynthesis, conversions
1.7.12 amino sugar conversions
1.7.13 amino acid conversion
1.7.14 polyamine biosynthesis
1.7.15 2'-deoxyribonucleotide metabolism
1.7.17 formyl-tetrahydrofolate biosynthesis
1.7.18 betaine biosynthesis
1.7.19 incorporation of metal ions
1.7.2 glyoxylate bypass
1.7.20 S-adenosyl methionine biosynthesis
1.7.21 glyoxylate degradation
1.7.22 carnitine metabolism
1.7.23 methylglyoxal metabolism
1.7.24 cyanate catabolism
1.7.25 glycolate metabolism
1.7.26 allantoin assimilation
1.7.27 pyridoxal 5'-phosphate salvage
1.7.28 pyruvate oxidation
1.7.29 acetate catabolism
1.7.3 pentose phosphate shunt, non-oxidative branch
1.7.30 threonine catabolism
1.7.31 aminobutyrate catabolism
1.7.32 putrescine catabolism
1.7.33 nucleotide and nucleoside conversions
1.7.34 peptidoglycan (murein) turnover, recycling
1.7.6 glycerol metabolism
1.7.7 galactose metabolism
1.7.8 gluconeogenesis
1.7.9 glucose metabolism
1.8 metabolism of other compounds
1.8.1 phosphorous metabolism
1.8.2 sulfur metabolism
1.8.3 nitrogen metabolism
2 information transfer
2.1 DNA related
2.1.1 DNA replication

2.1.2 DNA restriction/modification
2.1.3 DNA recombination
2.1.4 DNA repair
2.1.5 DNA degradation
2.2 RNA related
2.2.2 transcription related
2.2.3 RNA modification
2.2.4 RNA degradation
2.2.5 tRNA
2.2.6 rRNA, stable RNA
2.2.7 antisense RNA
2.3 protein related
2.3.1 amino acid -activation
2.3.2 translation
2.3.3 posttranslational modification
2.3.4 chaperoning, folding
2.3.5 export, signal peptide cleavage
2.3.6 turnover, degradation
2.3.7 nucleoproteins, basic proteins
2.3.8 ribosomal proteins
2.3.9 non-ribosomal peptide synthetase
3 regulation
3.1 type of regulation
3.1.1 DNA structure level
3.1.1.1 DNA bending, supercoiling, inversion
3.1.1.2 methylation
3.1.2 transcriptional level
3.1.2.1 sigma factors, anti-sigmafactors
3.1.2.2 activator
3.1.2.3 repressor
3.1.2.4 complex regulation
3.1.2.4.1 more than one signal needed
3.1.2.4.2 regulons or multilayer component regulatory systems
3.1.2.4.3 two component regulatory systems (external signal)
3.1.2.4.4 quorum sensing
3.1.2.5 action unknown
3.1.3 posttranscriptional
3.1.3.1 translation attenuation and efficiency
3.1.3.2 covalent modification, demodification, maturation
3.1.3.3 inhibition / activation of enzymes
3.1.3.4 proteases, cleavage of compounds
3.1.3.5 multilayer regulatory system
3.1.3.6 antisense RNA

3.1.4 regulation level unknown
3.3 genetic unit regulated
3.3.1 operon
3.3.2 regulon
3.3.3 stimulon
3.3.4 global
3.4 trigger
3.5 trigger modulation
4 transport
4.1 Channel-type Transporters
4.1.A alpha-type channels
4.1.A.1 The Voltage-gated Ion Channel (VIC) Superfamily
4.1.A.11 The Chloride Channel (ClC) Family
4.1.A.22 The Large Conductance Mechanosensitive Ion Channel (MscL) Family
4.1.A.23 The Small Conductance Mechanosensitive Ion Channel (MscS) Family
4.1.A.46 The H+- or Na+-translocating Bacterial Flagellar Motor (Mot) Family
4.1.A.8 The Major Intrinsic Protein (MIP) Family
4.1.B Beta barrel porins (The Outer Membrane Porin (OMP) Functional Superfamily)
4.1.B.1 The General Bacterial Porin (GBP) Family
4.1.B.10 The Nucleoside-specific Channel-forming Outer Membrane Porin (Tsx) Family
4.1.B.14 The Outer Membrane Receptor (OMR) Family
4.1.B.21 The OmpG Porin (OmpG) Family
4.1.B.3 The Sugar Porin (SP) Family
4.1.B.9 The FadL Outer Membrane Protein (FadL) Family
4.2 Electrochemical potential driven transporters
4.2.A Porters (Uni-, Sym- and Antiporters)
4.2.A.1 The Major Facilitator Superfamily (MFS)
4.2.A.10 The 2-Keto-3-Deoxygluconate Transporter (KDGT) Family
4.2.A.13 The C4-Dicarboxylate Uptake (Dcu) Family
4.2.A.14 The Lactate Permease (LctP) Family
4.2.A.15 The Betaine/Carnitine/Choline Transporter (BCCT) Family
4.2.A.16 The Telurite-resistance/Dicarboxylate Transporter (TDT) Family
4.2.A.17 The Proton-dependent Oligopeptide Transporter (POT) Family
4.2.A.19 The Ca2+:Cation Antiporter (CaCA) Family
4.2.A.2 The Glycoside-Pentoside-Hexuronide (GPH):Cation Symporter Family
4.2.A.20 The Inorganic Phosphate Transporter (PiT) Family
4.2.A.21 The Solute:Sodium Symporter (SSS) Family
4.2.A.23 The Dicarboxylate/Amino Acid:Cation (Na+ or H+) Symporter (DAACS) Family
4.2.A.25 The Alanine/Glycine:Cation symporter (AGCS) Family
4.2.A.26 The Branched Chain Amino Acid:Cation Symporter (LIVCS) Family
4.2.A.27 The Glutamate:Na+ Symporter (GltS) Family
4.2.A.3 The Amino Acid-Polyamine-Choline (APC) Family
4.2.A.33 The NhaA Na+:H+ Antiporter (NhaA) Family

4.2.A.34 The NhaB Na+:H+ Antiporter (NhaB) Family
4.2.A.36 The Monovalent Cation:Proton Antiporter-1 (CPA1) Family
4.2.A.37 The Monovalent Cation:Proton Antiporter-2 (CPA2) Family
4.2.A.38 The K+ Transporter (Trk) Family
4.2.A.39 The Nucleobase:Cation Symporter-1 (NCS1) Family
4.2.A.4 The Cation Diffusion Facilitator (CDF) Family
4.2.A.40 The Nucleobase:Cation Symporter-2 (NCS2) Family
4.2.A.41 The Concentrative Nucleoside Transporter (CNT) Family
4.2.A.42 The Hydroxy/Aromatic Amino Acid Permease (HAAAP) Family
4.2.A.44 The Formate-Nitrite Transporter (FNT) Family
4.2.A.45 The Metal Ion Transporter (MIT) Family
4.2.A.47 The Divalent Anion:Na+ Symporter (DASS) Family
4.2.A.49 The Ammonium Transporter (Amt) Family
4.2.A.53 The Sulfate Permease (SulP) Family
4.2.A.55 The Manganese (Nramp) Fam.
4.2.A.58 The Phosphate:Na+ Symporter (PNaS) Family
4.2.A.6 The Resistance-Nodulation-Cell Division (RND) Superfamily
4.2.A.61 The C4-dicarboxylate Uptake C (DcuC) Family
4.2.A.64 The Type V Secretory Pathway or Twin Arginine Targeting (Tat) Family
4.2.A.66 The Multi Antimicrobial Extrusion (MATE) Family
4.2.A.68 The p-Aminobenzoyl-glutamate Transporter (AbgT) Family
4.2.A.69 The Auxin Efflux Carrier (AEC) Family
4.2.A.7 The Drug/Metabolite Transporter (DMT) Superfamily
4.2.A.72 The K+ uptake permease (KUP) Family
4.2.A.75 The L-lysine exporter (LysE) Family
4.2.A.76 The Resistance to Homoserine/Threonine (RhtB) Family
4.2.A.8 The Gluconate:H+ Symporter (GntP) Family
4.2.A.9 The L-Rhamnose Transporter (RhaT) Family
4.2.C Ion-gradient-driven energizers
4.2.C.1 TonB Family of Auxiliary Proteins for Energization of OMR-mediated Transport
4.3 Primary Active Transporters
4.3.A Pyrophosphate Bond (ATP; GTP; P2) Hydrolysis-driven Active Transporters
4.3.A.1 The ATP-binding Cassette (ABC) Superfamily + ABC-type Uptake Permeases
4.3.A.1.a ABC superfamily ATP binding cytoplasmic component
4.3.A.1.am ABC superfamily, ATP binding and membrane component
4.3.A.1.m ABC superfamily, membrane component
4.3.A.1.p ABC superfamily, periplasmic binding component
4.3.A.2 The H+/Na+-translocating F-, V- and A-type ATPase (F-ATPase) Superfamily
4.3.A.3 The P-type ATPase (P-ATPase) Superfamily
4.3.A.4 The Arsenite-Antimonite (Ars) Efflux Family
4.3.A.5 The Type II (General) Secretory Pathway (IISP) Family
4.3.D Oxidoreduction-driven Active Transporters
4.3.D.1 The Proton- or sodium ion-translocating NADH Dehydrogenase (NDH) Family

4.3.D.4 The Proton-translocating Cytochrome Oxidase (COX) Superfamily
4.4 Group Translocators
4.4.A Phosphotransferase Systems (PEP-dependent PTS)
4.4.A.1 The PTS Glucose-Glucoside (Glc) Family
4.4.A.2 The PTS Fructose-Mannitol (Fru) Family
4.4.A.3 The PTS Lactose-N,N'-Diacetylchitobiose-betaucoside (Lac) Family
4.4.A.4 The PTS Glucitol (Gut) Family
4.4.A.5 The PTS Galactitol (Gat) Family
4.4.A.6 The PTS Mannose-Fructose-Sorbose (Man) Family
4.8.A Accessory Factors Involved in Transport
4.8.A.1 The Membrane Fusion Protein (MFP) Family
4.8.A.3 MPA1 Family auxillary transport protein
4.8.A.7 The Phosphotransferase System Enzyme I (EI) Family
4.8.A.8 The Phosphotransferase System HPr (HPr) Family
4.9 Transporters of Unknown Classification
4.9.A recognized transporters of unknown biochemical mechanism
4.9.A.1 The Polysaccharide Transporter (PST) Family
4.9.A.13 The Short Chain Fatty Acid Transporter (scFAT) Family
4.9.A.16 The Septal DNA Translocator (SDT) Family
4.9.A.17 The Metal Ion Transporter (MIT) Family
4.9.A.4 The Nicotinamide Mononucleotide (NMN) Uptake Permease (PnuC) Family
4.9.A.8 The Ferrous Iron Uptake (FeoB) Family
4.9.B Putative uncharacterized transport protein
4.9.B.10 The 6TMS Putative MarC Transporter (MarC) Family
4.9.B.18 The SecDF-associated Single Transmembrane Protein (SSTP) Family
4.9.B.21 The Frataxin (Frataxin) Family
4.9.B.22 The Putative Permease (PerM) Family
4.9.B.24 The Testis-Enhanced Gene Transfer (TEGT) Family
4.9.B.25 The YbbM (YbbM) Family
4.9.B.27 The YdjX-Z (YdjX-Z) Family
4.9.B.28 The YqaE (YqaE) Family
4.9.B.29 The YebN (YebN) Family
4.9.B.3 The Putative Bacterial Murein Precursor Exporter (MPE) Family
4.9.B.31 The YqiH (YqiH) Family
4.9.B.32 The Putative Vectorial Glycosyl Polymerization (VGP) Family
4.9.B.4 The Putative Efflux Transporter (PET) Family
4.9.B.6 The Toxic Hok/Gef Protein (Hok/Gef) Family
4.S substrate
5 cell processes
5.1 cell division
5.10 defense/survivial
5.11 DNA uptake
5.2 cell cycle physiology

5.3 motility (incl. chemotaxis, energytaxis, aerotaxis, redoxtaxis)
5.4 genetic exchange, recombination
5.5 adaptation to stress
5.5.1 osmotic pressure
5.5.2 temperature extremes
5.5.3 starvation response
5.5.4 pH response
5.5.5 dessication
5.5.6 other (mechanical, nutritional, oxidative stress)
5.5.7 Fe aquisition
5.6 protection
5.6.1 radiation
5.6.2 detoxification (xenobiotic metabolism)
5.6.3 cell killing
5.6.4 drug resistance/sensitivity
5.8 SOS response
6 cell structure
6.1 membrane
6.2 peptidoglycan (murein)
6.3 surface antigens (ECA, O antigen of LPS)
6.4 flagellum
6.5 pilus
6.6 ribosome
6.7 capsule (M and K antigens)
7 location of gene products
7.1 cytoplasm
7.2 periplasmic space
7.3 inner membrane
7.4 outer membrane
7.5 extracellular
8 extrachromosomal
8.1 prophage genes and phage related functions
8.2 plasmid related
8.3 transposon related
8.4 colicin related
9 DNA sites
10 cryptic genes
11 pathogenic related genes
12 Conserved Hypothetical
13 Hypothetical

References

Abd,H., Johansson,T., Golovliov,I., Sandström,G. and Forsman,M. (2003) Survival and growth of *Francisella tularensis* in *Acanthamoeba castellanii. App.Environ.Microbiol.* **69**, 600-606.

Abdul-Tehrani,H., Hudson,A.J., Chang,Y.S., Timms,A.R., Hawkins,C., Williams,J.M., Harrison,P.M., Guest,J.R. and Andrews,S.C. (1999) Ferritin mutants of *Escherichia coli* are iron deficient and growth impaired, and fur mutants are iron deficient. *J. Bacteriol.* **181**, 1415-1428.

Adams,M.A. and Jia,Z. (2005) Structural and biochemical evidence for an enzymatic quinone redox cycle in *Escherichia coli*: identification of a novel quinol monooxygenase. *J. Biol. Chem.* **280**, 8358-8363.

Alamuri,P., Mehta,N., Burke,A. and Maier,R.J. (2006) Regulation of the *Helicobacter pylori* Fe-S cluster synthesis protein NifS by iron, oxidative stress conditions and *fur*. *J. Bacteriol.* **188**, 5325-5330.

Alibek, K. (1999) *Biohazard: the chilling true story of the largest covert biological weapons program in the world, told from the inside by the man who ran it.* New York: Random House.

Allen,L.-A.H. (2003) Mechanisms of pathogenesis: evasion of killing by polynuclear leukocytes. *Microb. Infect.* **5**, 1329-1335.

Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410.

Ambion Inc. (2003) Saponin mediated lysis of eukaryotic cells harbouring bacterial pathogens. TechNotes. 10(1), online.

Andrews,S.C., Smith,J.M., Hawkins,C., Williams,J.M., Harrison,P.M. and Guest,J.R. (1993) Overproduction, purification and characterization of the bacterioferritin of *Escherichia coli* and a C-terminally extended variant. *Eur. J. Biochem.* **213**, 329-338.

Ansaruzzaman,M., Albert,M.J., Kühn,I., Faruque,S.M., Siddique,A.K. and Möllby,R. (1996) Differentiation of *Vibrio cholerae* O1 isolates with biochemical fingerprinting and comparison with ribotyping. *J. Diarrhoeal Dis. Res.* **14**, 248-254.

Anthony,L.D., Burke,R.D. and Nano,F.E. (1991) Growth of *Francisella* spp. in rodent macrophages. *Infect. Immun.* **59**, 3291-3296.

Arsène,F., Tomoyasu,T. and Bukau,B. (2000) The heat shock response of *Escherichia coli*. *Int. J. Food Microbiol.* **55**, 3-9.

Atkins,H.S., Dassa,E., Walker,N.J., Griffin,K.F., Harland,D.N., Taylor,R.R., Duffield,M.L. and Titball,R.W. (2006) The identification and evaluation of ATP binding cassette systems in the intracellular bacterium *Francisella tularensis*. *Res. Microbiol.* **157**, 593-604.

Atlas, R. M. (1993) *Handbook of Microbiological Media*. CRC Press.

Avashia,S.B., Petersen,J.M., Lindley,C.M., Schriefer,M.E., Gage,K.L., Cetron,M., DeMarcus,T.A., Kim,D.K., Buck,J., Montenieri,J.A., Lowell,J.L., Antolin,M.F., Kosoy,M.Y., Carter,L.G., Chu,M.C., Hendriks,K.A., Dennis,D.T. and Kool,J.L. (2004) First reported prairie dog-to-human tularemia transmission, Texas, 2002. *Emerg. Infect. Dis.* **10**, 483-486.

Bachmayer,H., Piette,L.H., Yasunobu,K.T. and Whitely,H.R. (1967) The binding sites of iron in rubredoxin from *Micrococcus aerogenes*. *Proc Natl Acad Sci U S A.* **57**, 122-127.

Bagg,A. and Neilands,J.B. (1987) Molecular mechanism of regulation of siderophore-mediated iron assimilation. *Microbiol. Rev.* **51**, 509-518.

Bakshi,C.S., Malik,M., Regan,K., Melendez,J.A., Metzger,D.W., Pavlov,V.M. and Sellati,T.J. (2006) Superoxide dismutase B gene (*sodB*)-deficient mutants of *Francisella tularensis* demonstrate hypersensitivity to oxidative stress and attenuated virulence. *J. Bacteriol.* **188**, 6443-6448.

Bandyopadhyay,S. and Cookson,M.R. (2004) Evolutionary and functional relationships within the DJ1 superfamily. *BMC Evol. Biol.* **4**.

Barbé,J., Villaverde,A. and Guerrero,R. (1987) Induction of the SOS response by hydroxyurea in *Escherichia coli* K12. *Mutat. Res.* **192**, 105-108.

Barker,J.H., Weiss,J., Apicella,M.A. and Nauseef,W.M. (2006) Basis for the failure of *Francisella tularensis* lipopolysaccharide to prime human polymorphonuclear leukocytes. *Infect. Immun.* **74**, 3277-3284.

Baron,G.S., Reilly,T.J. and Nano,F.E. (1999) The respiratory burst-inhibiting acid phosphatase AcpA is not essential for the intramacrophage growth or virulence of *Francisella novicida*. *FEMS Microbiol. Lett.* **176**, 85-90.

Beall,B. and Lutkenhaus,J. (1992) Impaired cell division and sporulation of a *Bacillus subtilis* strain with the *ftsA* gene deleted. *J. Bacteriol.* **174**, 2398-2403.

Beckman,J.S., Beckman,T.W., Chen,J., Marshall,P.A. and Freeman,B.A. (1990) Apparent hydroxyl radical production by peroxynitrite: implications for endothelial cell injury from nitric oxide and superoxide. *Proc. Natl. Acad. Sci. USA* **87**, 1620-1624.

Berdal,B.P., Mehl,R., Meidell,N.K., Lorentzen-Styr,A.M. and Scheel,O. (1996) Field investigations of tularemia in Norway. *FEMS Immun. Med. Microbiol.* **13**, 191-195.

Bernstein,J.A., Khodursky,A.B., Lin,P.H., Lin-Chao,S. and Cohen,S.N. (2002) Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc Natl Acad Sci U S A.* **99** , 9697-9702.

Bieniossek,C., Schalch,T., Bumann,M., Meister,M., Meier,R. and Bauman,U. (2006) The molecular architecture of the metalloprotease FtsH. *Proc Natl Acad Sci U S A.* **103**, 3066-3071.

Bischof,R. and Rogers,D.G. (2005) Serologic survey of select infectious diseases in coyotes and raccoons in Nebraska. *J. Wildl. Dis.* **41**, 787-791.

Bleuel,C., Groβe,C., Taudte,N., Scherer,J., Wesenberg,D., Krauβ,G.J., Nies,D.H. and Grass,G. (2005) TolC is involved in enterobactin efflux across the outer membrane of *Escherichia coli. J. Bacteriol.* **187**, 6701-6707.

Blick,R.J., Revel,A.T. and Hansen,E.J. (2003) FindGDPs: identification of primers for labeling microbial transcriptomes for DNA microarray analysis. *Bioinformatics* **19**, 1718-1719.

Bochner,B.R., Lee,P.C., Wilson,S.W., Cutler,C.W. and Ames,B.N. (1984) AppppA and related adenylated nucleotides are synthesized as a consequence of oxidation stress. *Cell* **37**, 225-232.

Boyce,J.D., Wilkie,I., Harper,M., Paustian,M.L., Kapur,V. and Adler,B. (2002) Genomic scale analysis of *Pasturella multocida* gene expression during growth within the natural chicken host. *Infect. Immun.* **70**, 6871-6879.

Braun,M. and Silhavy,T.J. (2002) Imp/OstA is required for cell envelope biogenesis in *Escherichia coli*. *Mol. Microbiol.* **45**, 1289-1302.

Braun,V., Hantke,K. and Koster,W. (1998) Bacterial iron transport: mechanisms, genetics, and regulation. *Met. Ions Biol. Syst.* **35**, 67-145.

Brittnacher, M., Rohmer, L., Zhou, Y., Abmayr, S., D'Argenio, D., Bovee, D., Chang, J., Chen, J., Drees, B., Ernst, R., Fong, C., Forsman, M., Gallagher, L., Gallis, B., Gillett, W., Goodlett, D., Guina, T., Guenthner, D., Haugen, E., Hayden, H., Jacobs, M., Kang, A., Larson Freeman, T., Levy, R., Lim, R., Manoil, C., Olson, M. V., Radey, M., Shaffer, S., Svensson, K., Taylor, G., Wasnick, M., Kaul, R. and Miller, S. I. (2006) *Francisella tularensis* subsp. *novicida* U112, complete genome.

Broekhuijsen,M., Larsson,P., Johansson,A., Bystrom,M., Eriksson,U., Larsson,E., Prior,R.G., Sjöstedt,A., Titball,R.W. and Forsman,M. (2003) Genome-wide DNA microarray analysis of *Francisella tularensis* strains demonstrates extensive genetic conservation within the species but identifies regions that are unique to the highly virulent *F. tularensis* subsp. *tularensis*. *J. Clinical Microbiol.* **41**, 2924-2931.

Brown,L. and Elliott,T. (1996) Efficient translation of the RpoS sigma factor in *Salmonella typhimurium* requires host factor 1, an RNA-binding protein encoded by the *hfq* gene. *J Bact* **178**, 3763-3770.

Burke,D.S. (1977) Immunization against tularemia: analysis of the effectiveness of live *Francisella tularensis* vaccine in prevention of laboratory-acquired tularemia. *J. Infect. Dis.* **135**, 55-60.

Butcher, P. D., Mangan, J. A. and Monahan, I. M. (1998) Intracellular gene expression: analysis of RNA from mycobacteria in macrophages using RT-PCR. In *Methods in Molecular Biology* ed. Parish,T. and Stoker,N.G. pp. 285-306. New Jersey, USA: Human Press.

Byström,M., Böcher,S., Magnusson,A., Praq,J. and Johansson,A. (2005) Tularemia in Denmark: identification of a *Francisella tularensis* subsp. *holarctica* strain by real-time PCR and high-resolution typing by multiple-locus variable-number tandem repeat analysis. *J. Clin. Microbiol.* **43**, 5355-5358.

Cabiscol,E., Tamarit,J. and Ros,J. (2000) Oxidative stress in bacteria and protein damage by reactive oxygen species. *Int. Microbiol.* **3**, 3-8.

Cammack,R., Johnson,C.E., Hall,D.O. and Rao,K.K. (1971) The state of iron atoms in hydroxylase iron-sulphur proteins. *Biochem. J.* **125**, 18P-19P.

Carroll,J.A., Dorward,D.W. and Gherardini,F.C. (1996) Identification of a transferrin-binding protein from *Borrelia burgdorferi* . *Infect. Immun.* **64**, 2911-2916.

Carver,F.J. and Frieden,E. (1978) Factors affecting the adenosine triphosphate induced release of iron from transferrin. *Biochem.* **17**, 167-172.

Chain,P., Larimer,F., Land,M., Stilwagen,S., Larsson,P., Bearden,S.W., Chu,M.C., Oyston,P.C.F., Forsman,M., Andersson,S., Lindler,L., Titball,R.W. and Garcia,E. (2006)

Complete genome sequence of *Francisella tularensis* LVS (Live Vaccine Strain). *Unpublished.*

Chamberlain,R.E. (1965) Evaluation of live tularemia vaccine prepared in a chemically defined medium. *App. Microbiol.* **13**, 232-235.

Chang,H.Y., Lee,J.H., Deng,W.L., Fu,T.F. and Peng,H.L. (1996) Virulence and outer membrane properties of a *galU* mutant of *Klebsiella pneumonia* CG43. *Microb. Pathogen.* **20**, 255-261.

Chaput,C., Labigne,A. and Boneca,I.G. (2007) Characterization of *Helicobacter pylori* lytic transglycosylases Slt and MltD. *J. Bacteriol.* **189**, 422-429.

Chaudhuri,R.R., Ren,C.P., Desmond,L., Vincent,G.A., Silman,N.J., Brehm,J.K., Elmore,M.J., Hudson,M.J., Forsman,M., Isherwood,K.E., Gurycová,D., Minton,N.P., Titball,R.W., Pallen,M.J. and Vipond,R. (2007) Genome sequencing shows that European isolates of *Francisella tularensis* subspecies *tularensis* are almost identical to US laboratory strain Schu S4. *PLoS ONE* **2**, e352.

Checroun,C., Wehrly,T.D., Fischer,E.R., Hayes,S.F. and Celli,J. (2006) Autophagy-mediated reentry of *Francisella tularensis* into the endocytic compartment after cytoplasmic replication. *PNAS* **103**, 14578-14583.

Cheng,Z.F. and Deutscher,M.P. (2002) Purification and characterization of the *Escherichia coli* exoribonuclease RNase R. Comparison with RNase II. *J. Biol. Chem.* **277**, 21624-21629.

Chitambar,C.R. and Narasimhan,J. (1991) Targeting iron-dependent DNA synthesis with gallium and transferrin-gallium. *Pathobiol.* **59**, 3-10.

Chiu,H.C., Lin,T.L. and Wang,J.T. (2007) Identification and characterization of organic solvent tolerance gene in *Helicobacter pylori*. *Helicobacter* **12**, 74-81.

Chou,J.H., Greenberg,J.T. and Demple,B. (1993) Posttranscriptional repression of *Escherichia coli* OmpF protein in response to redox stress: positive control of the *micF* antisense RNA by the *soxRS* locus. *J. Bacteriol.* **175**, 1026-1031.

Christiansen,J., Larsen,M.H., Ingmer,H., Søgaard-Andersen,L. and Kallipolitis,B.H. (2004) The RNA-binding protein Hfq of *Listeria monocytogenes*: role in stress tolerance and virulence. *J Bact* **186**, 3355-3362.

Christopher,G.W., Cieslak,T.J., Pavlin,J.A. and Eitzen,E.M.J. (1997) Biological warfare. A historical perspective. *JAMA* **278**, 412-417.

Clark,R.A. (1990) The human neutrophil respiratory burst. *J. Infect. Dis.* **161**, 1140-1147.

Clavel,T., Lazzaroni,J.C., Vianney,A. and Portalier,R. (1996) Expression of the *tolQRA* genes of *Escherichia coli* K-12 is controlled by the RcsC sensor protein involved in capsule synthesis. *Mol. Microbiol.* **19**, 19-25.

Clemens,D.L. and Horwitz,M.A. (1995) Characterization of the *Mycobacterium tuberculosis* phagosome and evidence that phagosomal maturation in inhibited. *J. Exp. Med.* **181**, 257-270.

Clemens,D.L., Lee,B.-Y. and Horwitz,M.A. (2004) Virulent and avirulent strains of *Francisella tularensis* prevent acidification and maturation of their phagosomes and escape into the cytoplasm in human macrophages. *Infect. Immun.* **72**, 3204-3217.

Clemens, D. L., Lee, Y. and Horwitz, M. A. (2005) *Francisella tularensis* enters macrophages via a novel process involving pseudopod loops. *Infect. Immun.* **73**, 5892-5902.

Cohen,S.P., McMurry,L.M. and Levy,S.B. (1988) *marA* locus causes decreased expression of OmpF porin in multiple-antibiotic-resistant (Mar) mutants of *Escherichia coli. J. Bacteriol.* **170**, 5416-5422.

Conlan,J.W. and North,R.J. (1992) Early pathogenesis of infection in the liver with the facultative intracellular bacteria *Listeria monocytogenes*, *Francisella tularensis*, and *Salmonella typhimurium* involves lysis of infected hepatocytes by leukocytes. *Infect. Immun.* **60**, 5164-5171.

Cox,J.M., Clayton,C.L., Tomita,T., Wallace,D.M., Robinson,P.A. and Crabtree,J.E. (2001) cDNA array analysis of *cag* pathogenicity island-associated *Helicobacter pylori* epithelial cell response genes. *Infect. Immun.* **69**, 6970-6980.

Croddy,E. and Krcalova,S. (2001) Tularemia, biological warfare and the battle for Stalingrad (1942-1943). *Mil. Med.* **166**, 837-838.

Curnutte,J.T. and Babior,B.M. (1974) Biological defense mechanisms. The effect of bacteria and serum on superoxide production by granulocytes. *J. Clin. Invest.* **53**, 1662-1672.

Dahlstrand,S., Ringertz,O. and Zetterberg,B. (1971) Airborne tularemia in Sweden. *Scand. J. Infect. Dis.* **3**, 7-16.

de Chastellier,C. and Berche,P. (1994) Fate of *Listeria monocytogenes* in murine macrophages: evidence for simultaneous killing and survival of extracellular bacteria. *Infect. Immun.* **62**, 543-553.

de Crécy-Lagard,V., El Yacoubi,B., de la Garza,R.D., Noiriel,A. and Hanson,A.D. (2007) Comparative genomics of bacterial and plant folate synthesis and salvage: predictions and validations. *Methods in enzymology* **425**, 153-183.

De Koning-Ward,T.F. and Robins-Browne,R.M. (1995) Contribution of urease to acid tolerance in *Yersinia enterocolitica*. *Infect. Immun.* **63**, 3790-3795.

Dean,C.R. and Goldberg,J.B. (2002) *Pseudomonas aeruginosa galU* is required for a complete lipopolysaccharide core and repairs a secondary mutation in a PA103 (serogroup O11) *wbpM* mutant. *FEMS Microbiol. Lett.* **210**, 277-283.

Dearlove,A.M. (2002) High throughput genotyping technologies. *Brief Funct. Genomic Proteomic* **1**, 139-150.

DebRoy,S., Dao,J., Söderberg,M., Rossier,O. and Cianciotto,N.P. (2006) *Legionella pneumophila* type II secretome reveals unique exoproteins and a chitinase that promotes bacterial persistence in the lung. *Proc Natl Acad Sci U S A.* **103**, 19146-19151.

Delcher,A.L., Harmon,D., Kasif,F., White,O. and Salzberg,S.L. (1999) Improved microbial gene identification with Glimmer. *Nuc. Acid Res.* **27**, 4636-4641.

Delihas,N. and Forst,S. (2001) MicF: an antisense RNA gene involved in response of *Escherichia coli* to global stress factors. *J. Mol. Biol.* **313**, 1-12.

Deng,K., Blick,R.J., Liu,W. and Hansen,E. (2006) Identification of *Francisella tularensis* genes affected by iron limitation. *Infect. Immun.* **74**, 4224-4236.

Dennis,D.T. (2001) Tularemia as a biological weapon. *JAMA* **285**, 2763-2773.

DeRosa,M.C., Sancar,A. and Barton,J.K. (2005) Electrically monitoring DNA repair by photolyase. *Proc Natl Acad Sci U S A.* **102**, 10788-10792.

Ding,Y., Davis,B.M. and Waldor,M.K. (2004) Hfq is essential for *Vibrio cholerae* virulence and downregulates sigma expression. *Mol. Microbiol.* **53**, 345-354.

Djaman,O., Outten,F.W. and Imlay,J.A. (2004) Repair of oxidised iron-sulfur clusters in *Escherichia coli*. *J. Biol. Chem.* **279**, 44590-44599.

Dorn,B.R., Dunn,W.A. and Progulske-Fox,A. (2002) Bacterial interactions with the autophagic pathway. *Cell. Microbiol.* **4**, 1-10.

Dorofe'ev,K.A. (1947) Classification of the causative agent of tularemia. *Symp. Res. Works Inst. Epidemiol. Mikobiol. Chita* **1**, 170-180.

Dorrell,N., Hinchliffe,S. and Wren,B.W. (2005) Comparative phylogenomics of pathogenic bacteria by microarray analysis. *Curr. Opin. Microbiol.* **8**, 620-626.

Dougherty,M.J. and Downs,D.M. (2006) A connection between iron-sulfur cluster metabolism and the biosynthesis of 4-amino-5-hydroxymethyl-2-methylpyrimidine pyrophosphate in *Salmonella enterica*. *Microbiol.* **152**, 2345-2353.

Dubuisson,J.F., Vianney,A., Hugouvieux-Cotte-Pattat,N. and Lazzaroni,J.C. (2005) Tol-Pal proteins are critical cell envelope components of *Erwinia chrysanthemi* affecting cell morphology and virulence. *Microbiol.* **151**, 3337-3347.

Dziejman,M., Balon,E., Boyd,D., Fraser,C.M., heidelberg,J.F. and Mekalanos,J.J. (2002) Comparative genomic analysis of *Vibrio cholerae*: Genes that correlate with cholera endemic and pandemic disease. *Proc.Nat.Acad.Sci., USA* **99**, 1556-1561.

Easton,A.M. and Kushner,S.R. (1983) Transcription of the *uvrD* gene of *Escherichia coli* is controlled by the *lexA* repressor and by attenuation. *Nuc. Acid. Res.* **11**, 8625-8640.

Eigelsbach,H.T. and Downs,C.M. (1962) Prophylactic effectiveness of live and killed tularemia vaccines. I. Production of vaccine and evaluation in the white mouse and guinea pig. *J Bacteriol.* **84**, 1020-1027.

Ellis,J., Oyston,P.C.F., Green,M. and Titball,R.W. (2002) Tularemia. *Clin. Microbiol. Rev.* **15**, 631-646.

Engel,H., Kazemier,B. and Keck,W. (1991) Murein-metabolizing enzymes from *Escherichia coli*: sequence analysis and controlled overexpression of the *slt* gene, which encodes the soluble lytic transglycosylase. *J. Bacteriol.* **173**, 6773-6782.

Ericsson,M., Tarnvik,A., Kuoppa,K., Sandstrom,G. and Sjostedt,A. (1994) Increased synthesis of DnaK, GroEL, and GroES homologs by *Francisella tularensis* LVS in response to heat and hydrogen peroxide. *Infect. Immun.* **62**, 178-183.

Eriksson,S., Lucchini,S., Thompson,A., Rhen,M. and Hinton,J.C.D. (2003) Unravelling the biology of macrophage infection by gene expression profiling of intracellular *Salmonella enterica. Molecular Microbiology* **47**, 103-118.

Ernst,F.D., Bereswill,S., Waidner,B., Stoof,J., Mader,U., Kusters,J.G., Kuipers,E.J., Kist,M., Vliet,A.H.M. and Homuth,G. (2005) Transcriptional profiling of *Helicobacter pylori* Fur- and iron-regulated gene expression. *Microbiol.* **151**, 533-546.

Eschenlauer,S.C., Coombs,G.H. and Mottram,J.C. (2006) PFPI-like genes are expressed in *Leishmania major* but are pseudogenes in other *Leishmania* species. *FEMS Microbiol. Lett.* **260**, 47-54.

Escolar,L., Pérez-Martín,J. and Lorenzo de,V. (1999) Opening the iron box: transcriptional metalloregulation by the Fur protein. *J Bacteriol.* **181**, 6223-6229.

Euzéby,J.P. (1997) List of bacterial names with standing in nomenclature: a folder available on the internet. *Int. J. Syst. Bacteriol.* **47**, 590-592.

Evans,M.E. (1985) *Francisella tularensis. Infect. Control* **6**, 381-383.

Evans,M.E., Gregory,D.W., Schaffner,W. and McGee,Z.A. (1985) Tularemia: a 30 year experience with 88 cases. *Medicine (Baltimore)* **64**, 251-269.

Farr,S.B. and Kogoma,T. (2006) Oxidative stress responses in *Escherichia coli* and *Salmonella typhimurium. Microbiol. Rev.* **55**, 561-585.

Fernandez-Recio,J., Walas,F., Federici,L., Venkatesh Pratap,J., Bavro,V.N., Miguel,R.N., Mizuguchi,K. and Luisi,B. (2004) A model of a transmembrane drug-efflux pump from Gram-negative bacteria. *FEBS Lett.* **578**, 5-9.

Fleischmann,R.D., Adams,M.D., White,O., Clayton,R.A., Kirkness,E.F., Kerlavage,A.R., Bult,C.J., Tomb,J.F., Dougherty,B.A., Merrick,J.M. and et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269** , 496-512.

Forestal,C.A., Malik,M., Catlett,S.V., Savitt,A.G., Benach,J.L., Sellati,T.J. and Furie,M.B. (2007) *Francisella tularensis* has a significant extracellular phase in infected mice. *J. Infect. Dis.* **196**, 134-137.

Forslund,A.L.K.K., Svensson,K., Salomonsson,E., Johansson,A., Byström,M., Oyston,P.C.F., Michell,S.L., Titball,R.W., Noppa,L., Frithz-Lindsten,E., Forsman,M. and Forsberg,A. (2006) Direct repeat-mediated deletion of a type IV pilin gene results in major virulence attenuation of *Francisella tularensis*. *Mol. Micro.* **59**, 1818-1830.

Forsman,M., Sandström,G. and Jaurin,B. (1990) Identification of *Francisella* species and discrimination of type A and type B strains of *F. tularensis* by 16S rRNA analysis. *Appl. Environ. Microbiol.* **56**, 949-955.

Forsman,M., Sandström,G. and Sjostedt,A. (1994) Analysis of 16S ribosomal DNA sequences of *Francisella* strains and utilization for determination of the phylogeny of the genus and for identification of strains by PCR. *Int. J. Syst. Bacteriol.* **44**, 38-46.

Fortier,A.H., Polsinelli,T., Green,S.J. and Nancy,C.A. (1992) Activation of macrophages for destruction of *Francisella tularensis*: Identification of cytokines, effector cells, and effector molecules. *Infect. Immun.* **60**, 817-825.

Fortier,A.H., Leiby,D.A., Narayanan,R.B., Asafoadjei,E., Crawford,R.M., Nacy,C.A. and Meltzer,M.S. (1995) Growth of *Francisella tularensis* LVS in macrophages: the acidic intracellular compartment provides essential iron required for growth. *Infect. Immun.* **63**, 1478-1483.

Froelich,J.M., Tran,K. and Wall,D. (2006) A *pmrA* constitutive mutant sensitizes *Escherichia coli* to deoxycholic acid. *J. Bacteriol.* **188**, 1180-1183.

Fucci,L., Oliver,C.N., Coon,M. and Stadtman,E.R. (1983) Inactivation of key metabolic enzymes by mixed function oxidation reactions: possible implication in protein turnover and ageing. *Proc. Natl. Acad. Sci. USA* **80**, 1521-1525.

Fulop,M., Webber,T. and Manchee,R. (1993) Activation of the complement system by *Francisella tularensis* lipopolysaccharide. *New Microbiol.* **16**, 141-147.

Gavrillin,M.A., Bouakl,I.J., Knatz,N.L., Duncan,M.D., Hall,M.W., Gùnn,J.S. and Wewers,M.D. (2006) Internalization and phagosome escape required for *Francisella* to induce human monocyte Il-1β processing and release. *PNAS* **103**, 141-146.

Gebran,S.J., Yamamoto,Y., Newton,C., Tomioka,M., Widen,R., Klein,T.W. and Friedman,H. (1995) LPS inhibits the intracellular growth of *Legionella pneumophila* in thioglycolate elicited murine peritoneal macrophages by iron-dependent, tryptophan independent, oxygen-independent, and arginine-independent mechanisms. *J. Leukoc. Biol.* **57**, 80-87.

Geissler,B. and Margolin,W. (2005) Evidence for functional overlap among multiple bacterial cell division proteins: compensating for the loss of FtsK. *Mol. Microbiol.* **58**, 596-612.

Georgopoulos,C. and Ang,D. (1990) The *Escherichia coli groE* chaperonins. *Semin. Cell Biol.* **1**, 19-25.

Ghim,S.Y. and Neuhard,J. (1994) The pyrimidine biosynthesis operon of the thermophile *Bacillus caldolyticus* includes genes for uracil phosphoriboslytransferase and uracil permease. *J. Bacteriol.* **176**, 3698-3707.

Ghosh,S. and Deutscher,M.P. (1999) Oligoribonuclease is an essential component of the mRNA decay pathway. *Proc Natl Acad Sci U S A.* **96**, 4372-4377.

Gibbons,H.S., Kalb,S.R., Cotter,R.J. and Raetz,C.R. (2005) Role of Mg^{2+} and pH in the modification of *Salmonella* lipid A after endocytosis by macrophage tumour cells. *Mol. Microbiol.* **55**, 425-440.

Golovliov,I., Ericsson,M., Sandström,G., Tärnvik,A. and Sjöstedt,A. (1997) Identification of proteins of *Francisella tularensis* induced during growth in macrophages and cloning of the gene encoding a prominently induced 23-kilodalton protein. *Infect. Immun.* **65**, 2183-2189.

Golovliov,I., Sjöstedt,A., Mokrievich,A. and Pavlov,V. (2003) A method for allelic replacement in *Francisella tularensis. FEMS Microbiol. Lett.* **222**, 273-280.

Gordon,J.R., McLaughlin,B.G. and Nitiuthai,S. (1983) Tularaemia transmitted by ticks (*Dermacentor andersoni*) in Saskatchewan. *Can. J. Comp. Med.* **47**, 408-411.

Grentzmann,G., Brechemier-Baey,D., Heurque,V., Mora,L. and Buckingham,R.H. (1994) Localization and characterization of the gene encoding release factor RF3 in *Escherichia coli. Proc Natl Acad Sci U S A.* **91**, 5848-5852.

Guckenberger,M., Kurz,S., Aepinus,C., Theiss,S., Haller,S., Leimbach,T., Panzner,U., Weber,J., .Hubert,P., Unkmeir,A., Frosch,M. and Dietrich,G. (2002) Analysis of the heat shock response of *Neisseria meningitidis* with cDNA- and oligonucleotide-based DNA microarrays. *J. Bacteriol.* **184**, 2546-2551.

Guerinot,M.L. (1994) Microbial iron transport. *Annu. Rev. Microbiol.* **48**, 743-772.

Hajjar,A.M., Harvey,M.D., Shaffer,S.A., Goodlett,D.R., Sjöstedt,A., Edebro,H., Forsman,M., Byström,M., Pelletier,M.J., Wilson,C.B., Miller,S.I., Skerrett,S.J. and

Ernst,R.K. (2006) Lack of *in vitro* and *in vivo* recognition of *Francisella tularensis* subspecies by lipopolysaccharide Toll-like receptors. *Infect. Immun.* **74**, 6730-6738.

Halio,S.B., Blumentals,I.I., Short,S.A., Merrill,B.M. and Kelly,R.M. (1996) Sequence, expression in *Escherichia coli*, and analysis of the gene encoding a novel intracellular protease (PfpI) from the hyperthermophilic archaeon *Pyrococcus furiosus*. *J Bact* **178**, 2605-2612.

Hall,M.N. and Silhavy,T.J. (1981) Genetic analysis of the major outer membrane proteins of *Escherichia coli*. *Annu. Rev. Genet.* **15**, 91-142.

Hanawa,T., Yamamoto,T. and Kamiya,S. (1995) *Listeria monocytogenes* can grow in macrophages without the aid of proteins induced by environmental stresses. *Infect. Immun.* **63**, 4595-4599.

Harris,S. (1992) Japanese biological warfare research on humans: a case study of microbiology and ethics. *Ann. N. Y. Acad. Sci.* **666**, 21-52.

Hartley, M. G. (2003) Personal communication regarding *Francisella*-infection of macrophages.

Hartley,M.G., Taylor,R., Prior,J.L., Newstead,S., Hitchen,P.G., Morris,H.R., Dell,A. and Titball,R.W. (2006) Grey variants of the live vaccine strain of *Francisella tularensis* lack lipopolysaccharide O-antigen, show reduced ability to survive in macrophages and do not induce protective immunity in mice. *Vaccine* **24**, 989-996.

Hecker,M., Schumann,W. and Völker,U. (1996) Heat-shock and general stress response in *Bacillus subtilis*. *Mol. Microbiol.* **19**, 417-428.

Heidelberg,J., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwinn,M.L., Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L., Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D., Ermolaeva,M.D., Vamathevan,J., Bass,S., Qin,H., Dragoi,I., Sellers,P., McDonald,L., Utterback,T.R., Fleischmann,R.D., Nierman,W., White,O., Salzberg,S.L., Smith,H.O., Colwell,R.R., Mekalanos,J.J., Venter,J.C. and Fraser,C.M. (2000) DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae* . *Nature* **406**, 477-483.

Hernychova,L., Stulik,J., Halada,P., Maceia,A., Kroca,M., Johansson,T. and Malina,M. (2001) Construction of a *Francisella tularensis* two-dimensional electrophoresis protein database. *Proteomics* **1**, 508-515.

Hiles,I.D., Powell,L.M. and Higgins,C.F. (1987) Peptide transport in *Salmonella typhimurium*: molecular cloning and characterization of the oligopeptide permease genes. *Mol Gen Genet* **206**, 101-109.

Hollis,D.G., Weaver,R.E., .Steigerwalt,A.G., Wenger,J.D., Moss,C.W. and Brenner,D.J. (1989) *Francisella philomiragia* comb. *nov.* (formerly *Yersinia philomiragia*) and *Francisella tularensis* biogroup *novicida* (formerly *Francisella novicida*) associated with human disease. *J. of Clinical Microbiol.* **27**, 1601-1608.

Hopla,C.E. (1974) The ecology of tularemia. *Adv. Vet. Sci. Comp. Med.* **18**, 25-53.

Hrstka,R., Stulík,J. and Vojtesek,B. (2005) The role of MAPK signal pathways during *Francisella tularensis* LVS infection-induced apoptosis in murine macrophages. *Microbes Infect.* **7**, 619-625.

Hubálek,M., Hernychová,L., Brychta,M., Lenco,J., Zechovská,J. and Stulík,J. (2004) Comparative proteome analysis of cellular proteins extracted from highly virulent

Francisella tularensis ssp. *tularensis* and less virulent *F. tularensis* ssp. *holarctica* and *F. tularensis* ssp. *mediaasiatica*. *Proteomics* **4**, 3048-3060.

Hunter, A. (2005) Personal communication regarding BlueFuse advice: e-mail from G. Snudden. graham.snudden@cambridgebluegnome.com.

Imlay,J.A. and Linn,S. (1986) Mutagenesis and stress responses induced in *Escherichia coli* by hydrogen peroxide. *J Bact* **169**, 2967-2976.

Ishibashi,Y. and Arai,T. (1995) *Salmonella typhi* does not inhibit phagosome-lysosome fusion in human monocyte derived-macrophages. *FEMS Immunol.Med.Microbiol.* **12**, 55-61.

Ishii,T., Yasuda,K., Akatsuka,A., Hino,O., Hartman,P.S. and Ishii,N. (2005) A mutation in the SDHC gene of complex II increases oxidative stress, resulting in apoptosis and tumorigenesis. *Cancer Res.* **65**, 203-209.

Isolar,N.R., Allman,S.L., Golovlev,V.V. and Chen,C.H. (2001) MALDI-TOF mass spectrometric method for detection of hybridized DNA oligomers. *Anal. Chem.* **73**, 2126-2131.

Izard,T. (2002) The crystal structures of phosphopantetheine adenylyltransferase with bound substrates reveal the enzymes catalytic mechanism. *J. Mol. Biol.* **315**, 487-495.

Jaffé,A. and D'Ari,R. (1985) Regulation of chromosome segregation in *Escherichia coli*. *Ann. Inst. Pasteur Microbiol.* **136A**, 159-164.

Johansson,A., Farlow,J., Larsson,P., Dukerich,M., Chambers,E., Byström,M., Fox,J., Chu,M.C., Forsman,M., Sjöstedt,A. and Keim,P. (2004) Worldwide genetic relationships

among *Francisella tularensis* isolates determined by multiple-locus variable-number tandem repeat analysis. *J. Bacteriol.* **186**, 5808-5818.

Jung,Y.S., Gao-Sheridan,H.S., Christiansen,J., Dean,D.R. and Burgess,B.K. (1999) Purification and biophysical characterization of a new [2Fe-2S] ferredoxin from *Azotobacter vinelandii*, a putative [Fe-S] cluster assembly/repair protein. *J. Biol. Chem.* **274**, 32402-32410.

Kagan,B.L., Selsted,M.E., Ganz,T. and Lehrer,R.I. (1990) Antimicrobial defensin peptides form voltage-dependent ion-permeable channels in planar lipid bilayer membranes. *Proc Natl Acad Sci U S A.* **87**, 210-214.

Kane,M.D., Jatkoe,T.A., Stumpf,C.R., Lu,J., Thomas,J.D. and Madore,S.J. (2000) Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. *Nuc. Acid. Res.* **28**, 4552-4557.

Kanemori,M., Yanaqi,H. and Yura,T. (1999) Marked instability of the sigma(32) heat shock transcription factor at high temperature. Implications for heat shock regulation. *J. Biol. Chem.* **274**, 22002-22007.

Kang,Y.H., Chung,S.J., Kang,I.J., Park,J.H. and Bünger,R. (2001) Intramitochondrial pyruvate attenuates hydrogen peroxide-induced apoptosis in bovine pulmonary artery endothelium. *Mol. Cell Biochem.* **216**, 37-46.

Kannangara,C.G., Vothknecht,U.C., Hansson,M. and von Wettstein,D. (1997) Magnesium chelatase: association with ribosomes and mutant complementation studies identify barley subunit Xantha-G as a functional counterpart of *Rhodobacter* subunit. *Mol. Gen. Genet.* **254**, 85-92.

Karaolis,D.K., Johnson,J.A., Bailey,C.C., Boedeker,E.C., Kaper,J.B. and Reeves,P.R. (1998) A *Vibrio cholerae* pathogenicity island associated with epidemic and pandemic strains. *Proc Natl Acad Sci U S A.* **95**, 3134-3139.

Kemp,J.T., Driks,A. and Losick,R. (2002) FtsA mutants of *Bacillus subtilis* impaired in sporulation. *J. Bacteriol.* **184**, 3856-3863.

Keyer,K., Gort,A.S. and Imlay,J.A. (1995) Superoxide and the production of oxidative DNA damage. *J Bacteriol.* **177**, 6782-6790.

Kidgell,C. and Winzeler,E.A. (2005) Elucidating genetic diversity with oligonucleotide arrays. *Chromosome Res.* **13**, 225-235.

Kim,N., Marcus,E.A., Wen,Y., Weeks,D.L., Scott,D.R., Jung,H.C., Song,I.S. and Sachs,G. (2004) Genes of *Helicobacter pylori* regulated by attachment to AGS cells. *Infect. Immun.* **72**, 2358-2368.

Kitagawa,M., Wada,C., Yoshioka,S. and Yura,T. (1991) Expression of ClpB, an analogue of the ATP-dependent protease regulatory subunit of *Escherichia coli*, is controlled by a heat shock sigma factor (sigma 32). *J. Bacteriol.* **173**, 4247-4253.

Kohler,S., Foulongne,V., Ouahrani-Bettache,S., Bourg,G., Teyssier,J., Ramuz,M. and Liautard,J.-P. (2002) The analysis of the intramacrophagic virulome of *Brucella suis* deciphers the environment encountered by the pathogen inside the macrophage host cell. *PNAS* **99**, 15711-15716.

Konopka,K. and Romslo,I. (1981) Studies on the mechanism of pyrophosphate-mediated uptake of iron from transferrin by isolated rat-liver mitochondria. *Eur. J. Biochem.* **117**, 239-244.

Korch,S.B. and Hill,T.M. (2006) Ectopic overexpression of wild-type and mutant *hipA* genes in *Escherichia coli*: effects on macromolecular synthesis and persister formation. *J. Bacteriol.* **188**, 3826-3836.

Krithika,R., Marathe,U., Saxena,P., Ansari,M.Z., Mohanty,D. and Gokhale,R.S. (2006) A genetic locus required for iron acquisition in *Mycobacterium tuberculosis. Proc Natl Acad Sci U S A.* **103**, 2069-2074.

Kwaik,Y.A., Eisenstein,B.I. and Engleberg,N.C. (1993) Phenotypic modulation by *Legionella pneumophila* upon infection of macrophages. *Infect. Immun.* **61**, 1320-1329.

Lai,X.H. and Sjöstedt,A. (2003) Delineation of the molecular mechanisms of *Francisella tularensis*-induced apoptosis in murine macrophages. *Infect. Immun.* **71**, 4642-4646.

Lai,X.H., Golovliov,I. and Sjöstedt,A. (2004) Expression of IglC is necessary for intracellular growth and induction of apoptosis in murine macrophages by *Francisella tularensis. Microb. Path.* **37**, 225-230.

Larson,C.L., Wicht,W. and Jellison,W.L. (1955) A new organism resembling *P. tularensis* isolated from water. *Public Health Rep.* **70**, 253-258.

Larsson,P., Oyston,P.C.F., Chain,P., Chu,M.C., Duffield,M., Fuxelius,H., Garcia,E., Halltorp,G., Johansson,D., Isherwood,K.E., Karp,P.D., Larsson,E., Liu,Y., Michell,S.L., Prior,J.L., Prior,R.G., Malfatti,S., Sjostedt,A., Svensson,K., Thompson,N., Vergez,L., Wagg,J.K., Wren,B.W., Lindler,L.E., Andersson,S.G.E., Forsman,M. and Titball,R.W. (2005) The complete genome sequence of *Francisella tularensis*, the causative agent of tularemia. *Nat. Genet.* **37**, 153-159.

Layer,G., Gaddam,S.A., Ayala-Castro,C.N., Ollagnier-de-Choudens,S., Lascoux,D., Fontecave,M. and Outten,F.W. (2007) SufE transfers sulfur from SufS to SufB for iron-sulfur cluster assembly. *J. Biol. Chem.* **282**, 13342-13350.

Lazdunski,C.J., Bouveret,E., Rigal,A., Journet,L., Lloubès,R. and Bénédetti,H. (1998) Colicin import into *Escherichia coli* cells. *J Bact* **180**, 4993-5002.

Lazzaroni,J.C., Germon,P., Ray,M.C. and Vianney,A. (1999) The Tol proteins of *Escherichia coli* and their involvement in the uptake of biomolecules and outer membrane stability. *FEMS Microbiol. Lett.* **177**, 191-197.

Le Brun,N.E., Andrews,S.C., Guest,J.R., Harrison,P.M., Moore,G.R. and Thomson,A.J. (1995) Identification of the ferroxidase centre of *Escherichia coli* bacterioferritin. *Biochem. J.* **312**, 385-392.

Lee,P.C., Bochner,B.R. and Ames,B.N. (1983) Diadenosine 5',5'''-P1,P4-tetraphosphate and related adenylated nucleotides in *Salmonella typhimurium*. *J. Biol. Chem.* **258**, 6827-6834.

Lee,S.J., Kim,S.J., Kim,I.K., Ko,J., Jeong,C.S., Kim,G.H., Park,C., Kang,S.O., Suh,P.G., Lee,H.S. and Cha,S.S. (2003) Crystal structures of human DJ-1 and *Escherichia coli* HSP31, which share an evolutionarily conserved domain. *J. Biol. Chem.* **278**, 44552-44559.

Lewin,T.M. and Webster,R.E. (1996) Membrane insertion characteristics of the various transmembrane domains of the *Escherichia coli* TolQ protein. *J. Biol. Chem.* **271**, 14143-14149.

Li, Ziyi and Demple B. (1994) SoxS, an activator of superoxide stress genes in *Escherichia coli. J. Biol. Chem.* **269**, 18371-18377.

Li,X.Z., Nikaido,H. and Poole,K. (1995) Role of *mexA-mexB-oprM* in antibiotic efflux in *Pseudomonas aeruginosa. Antimicrob. Agents Chemother.* **39**, 1948-1953.

Lindgren,H., Golovliov,I., Baranov,V., Ernst,R.E., Telepnev,M. and Sjöstedt,A. (2004) Factors affecting the escape of *Francisella tularensis* from the phagolysosome. *J. Med.Microbiol.* **53**, 953-958.

Lindgren,H., Stenmark,S., Chen,W., Tärnvik,A. and Sjöstedt,A. (2004) Distinct roles of reactive nitrogen and oxygen species to control infection with the facultative intracellular bacterium *Francisella tularensis. Infect. Immun.* **72**, 7172-7182.

Lindgren,H., Stenman,L., Tärnvik,A. and Sjöstedt,A. (2005) The contribution of reactive nitrogen and oxygen species to the killing of *Francisella tularensis* LVS by murine macrophages. *Microb. Infect.* **7**, 467-475.

Lindgren,H., Shen,H., Zingmark,C., Golovliov,I., Conlan,J.W. and Sjöstedt,A. (2007) Resistance of *Francisella tularensis* strains against reactive nitrogen and oxygen species with special reference to the role of KatG. *Infect. Immun.* **75**, 1303-1309.

Lockhart,D.J., Dong,H., Byrne,M.C., Follettie,M.T., Gallo,M.V., Chee,M.S., Mittmann,M., Wang,C., Kobayashi,M., Horton,H. and Brown,E.L. (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.* **14**, 1675-1680.

Loiseau,L., Ollagnier-de-Choudens,S., Nachin,L., Fontecave,M. and Barras,F. (2003) Biogenesis of Fe-S cluster by the bacterial Suf system: SufS and SufE for a new type of cysteine desulfarase. *J. Biol. Chem.* **278**, 38352-38359.

Löfgren,S., Tärnvik,A., Bloom,G.D. and Sjöberg,W. (1983) Phagocytosis and killing of *Francisella tularensis* by human polymorphonuclear leukocytes. *Infect. Immun.* **39**, 715-720.

Löfgren,S., Tärnvik,A., Thore,M. and Carlsson,J. (1984) A wild and attenuated strain of *Francisella tularensis* differ in susceptibility to hypochlorous acid: a possible explanation of their different handling by polymorphonuclear leukocytes. *Infect. Immun.* **43**, 730-734.

Lugtenberg,B., Peters,R., Bernheimer,H. and Berendsen,W. (1976) Influence of cultural conditions and mutations on the composition of the outer membrane proteins of *Escherichia coli. Mol Gen Genet* **147**, 251-262.

Lynch, A. S. and Lin, E. C. S. (1996) Responses to molecular oxygen. In *Escherichia coli and Salmonella: Cellular and Molecular Biology* ed. Neidhardt,F.C. and et al. pp. 1526-1538. ASM Press.

Ma,L.C., Lee,S.L., Lee,C.Z., Yin,S.J. and Wang,J.T. (2005) Molecular cloning and kinetic characterization of an aldehyde dehydrogenase gene in *Klebsiella pneumoniae*. *J. Formos. Med. Assoc.* **104**, 221-230.

Mandin,P., Fsihi,H., Dussurget,O., Vergassola,M., Milohanic,E., Toledo-Arana,A., Lasa,I., Johansson,J. and Cossart,P. (2005) VirR, a response regulator critical for *Listeria monocytogenes* virulence. *Mol. Microbiol.* **57**, 1367-1380.

Manganelli,R., Voskuil,M.I., Schoolnik,G.K. and Smith,I. (2001) The *Mycobacterium tuberculosis* ECF sigma factor sigmaE: role in global gene expression and survival in macrophages. *Mol. Microbiol.* **41**, 423-437.

Marchal,K., De Keersmaeckers,S., Monsieurs,P., van Boxel,N., Lemmens,K., Thijs,G., Vanderleyden,J. and De Moor,B. (2004) *In silico* identification and experimental validation of PmrAB targets in *Salmonella typhimurium* by regulatory motif detection. *Genome Biol.* **5**, R9.

Marshall,B.J., Barrett,L.J., Prakash,C., McCallum,R.W. and Guerrant,R.L. (1990) Urea protects *Helicobacter* (*Campylobacter*) *pylori* from the bactericidal effect of acid. *Gastroenterology* **99**, 697-702.

Martin,C. (1996) The *clp* (CS31A) operon is negatively controlled by Lrp, ClpB, and L-alanine at the transcriptional level. *Mol. Microbiol.* **21**, 281-292.

Martin,T., Holmes,I.H., Wobeser,G.A., Anthony,R.F. and Greefkes,I. (1982) Tularaemia in Canada with a focus on Saskatchewan. *Can. Med. Assoc. J.* **127**, 279-282.

Massé,E., Majdalani,N. and Gottesman,S. (2003) Regulatory roles for small RNAs in bacteria. *Curr. Opin. Microbiol.* **6**, 120-124.

McCaffrey,R.L. and Allen,L.-A.H. (2006) Pivotal advance: *Francisella tularensis* LVS evades killing by human neutrophils via inhibition of the respiratory burst and phagosome escape. *J. of Leukocyte Biology* **81**, 1-7.

McCormick,M.L., Buettner,G.R. and Britigan,B.E. (1998) Endogenous superoxide dismutase levels regulate iron-dependent hydroxyl radical formation in *Escherichia coli* exposed to hydrogen peroxide. *J Bact* **180**, 622-625.

McCoy,G.W. and Chapin,C.W. (1912) Further observations on a plague-like disease of rodents with a preliminary note on the causative agent, *Bacterium tularense*. *J. Infect. Dis.* **10**, 61-72.

McCrumb,F.R. (1961) Aerosol infection of man with *Pasturella tularensis* . *Bacteriol. Rev.* **25**, 262-267.

Mekalanos,J.J. (1992) Environmental Signals Controlling Expression of Virulence Determinants in Bacteria. *J. Bacteriol.* **174**, 1-7.

Merrell,C.L. and Wright,D.N. (1978) A serologic survey of mule deer and elk in Utah. *J. Wildl. Dis.* **14**, 471-478.

Merrell,D.S., Goodrich,M.L., Otto,G., Tompkins,L.S. and Falkow,S. (2003) pH-regulated gene expression of the gastric pathogen *Helicobacter pylori*. *Infect. Immun.* **71**, 3529-3539.

Mey,A.R., Wyckoff,E.E., Kanukurthy,V., Fisher,C.R. and Payne,S.M. (2005) Iron and fur regulation in *Vibrio cholerae* and the role of fur in virulence. *Infect. Immun.* **73**, 8167-8178.

Michaels,M.L. and Miller,J.H. (1992) The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-hydroxyguanine (7,8-dihydro-8-oxoguanine). *J. Bacteriol.* **174**, 6321-6325.

Mikalsen,J., Olsen,A.B., Tengs,T. and Colquhoun,D.J. (2007) *Francisella philomiragia* subsp. *noatunenesis* subsp. nov., isolated from farmed Atlantic cod (*Gadus morhua* L.). *Int. J. Syst. Evol. Microbiol.* **57**, 1960-1965.

Mikuni,O., Ito,K., Moffat,J., Matsumura,K., McCaughan,K., Nobukuni,T., Tate,W. and Kakamura,Y. (1994) Identification of the *prfC* gene, which encodes peptide-chain-release factor 3 of *Escherichia coli*. *Proc Natl Acad Sci U S A.* **91**, 5798-5802.

Miller,C., Thomsen,L.E., Gaggero,C., Mosseri,R., Ingmer,H. and Cohen,S.N. (2004) SOS response induction by beta-lactams and bacterial defense against antibiotic lethality. *Science* **305** , 1629-1631.

Miller,R.A. and Britigan,B.E. (1997) Role of oxidants in microbial pathophysiology. *Clin. Microbiol. reviews* **10**, 1-18.

Milne,L., Perrin,D.M. and Sigman,D.S. (2001) Oligoribonucleotide-based gene-specific transcription inhibitors that target the open complex. *Methods* **23**, 160-168.

Mio,T., Kokado,M., Arisawa,M. and Yamada-Okabe,H. (2000) Reduced virulence of *Candida albicans* mutants lacking the GNA1 gene encoding glucosamine-6-phosphate acetyltransferase. *Microbiol.* **146**, 1753-1758.

Moalem,S., Weinberg,E.D. and Percy,M.E. (2004) Hemochromatosis and the enigma of misplaced iron: implications for infectious disease and survival. *Biometals* **17**, 135-139.

Mockler,T.C., Chan,S., Sundaresan,A., Chen,H., Jacobsen,S.E. and Ecker,J.R. (2005) Applications of DNA tiling arrays for whole-genome analysis. *Genomics* **85**, 1-15.

Mohapatra,N.P., Balagopal,A., Soni,S., Schlesinger,L.S. and Gunn,J.S. (2007) AcpA is a *Francisella* acid phosphatase that affects intramacrophage survival and virulence. *Infect. Immun.* **75**, 390-396.

Morgan,E.H. (1979) Studies on the mechanism of iron release from transferrin. *Biochim. Biophys. Acta.* **580**, 312-326.

Morgan,R.W., Christman,M.F., Jacobson,F.S., Storz,G. and Ames,B.N. (1986) Hydrogen peroxide-inducible proteins in *Salmonella typhimurium* overlap with heat shock and other stress proteins. *Proc Natl Acad Sci U S A.* **83**, 8059-8063.

Moyed,H.S. and Bertrand,K.P. (1983) *hipA*, a newly recognized gene of *Escherichia coli* K-12 that affects frequency of persistence after inhibition of murein synthesis. *J. Bacteriol.* **155**, 768-775.

Mörner,T. (1992) The ecology of tularemia. *Rev. Sci. Tech. Off. Int. Epiz* **11**, 1130.

Mukherjee,S., Keitany,G., Li,Y., Wang,Y., Ball,H.L., Goldsmith,E.J. and Orth,K. (2006) *Yersinia* YopJ acetylates and inhibits kinase activation by blocking phosphorylation. *Science* **312**, 1211-1214.

Murray,P.J. and Young,R.A. (1992) Stress and immunological recognition in host-pathogen interactions. *J Bacteriol.* **174**, 4193-4196.

Mushegian,A.R., Fullner,K.J., Koonin,E.V. and Nester,E.W. (1996) A family of lysozyme-like virulence factors in bacterial pathogens of plants and animals. *Proc Natl Acad Sci U S A.* **93** , 7321-7326.

Nakamura,K., Yamaki,M., Sarada,M., Nakayama,S., Vibat,C.R., Gennis,R.B., Nakayashiki,T., Inokuchi,H., Kojima,S. and Kima,K. (1996) Two hydrophobic subunits are essential for the heme b ligation and functional assembly of complex II (succinate-ubiquinone oxidoreductase) from *Escherichia coli. J. Biol. Chem.* **271**, 521-527.

Nakao,H., Watanabe,H., Nakayama,S. and Takeda,T. (1995) *yst* gene expression in *Yersinia enterocolitica* is positively regulated by a chromosomal region that is highly homologous to *Escherichia coli* host factor 1 gene (*hfq*). *Mol. Microbiol.* **18**, 859-865.

Nano,F.E., Zhang,N., Cowley,S.C., Klose,K.E., Cheung,K.K.M., Roberts,M.J., Ludu,J.S., Letendre,G.W., Meierovics,A.I., Stephens,G. and Elkins,K.L. (2004) A

Francisella tularensis pathogenicity island required for intramacrophagic growth. *J. Bacteriol.* **186**, 6430-6436.

Nathan,C. and Xie,Q. (1994) Nitric oxide synthases: roles, tolls, and controls. *Cell* **78**, 915-918.

Neale,F.C. (1955) The demonstration of the iron-binding globulin (transferrin) in serum and urine proteins by use of ^{59}Fe combined with paper electrophoresis. *J. Clin. Pathol.* **8**, 334-337.

Neilands,J.B. (1995) Siderophores: structure and function of microbial iron transport compounds. *J. Biol. Chem.* **270**, 26723-26726.

Nesper,J., Lauriano,C.M., Klose,K.E., Kapfhammer,D., Kraiss,A. and Reidl,J. (2001) Characterization of *Vibrio cholerae* O1 El tor *galU* and *galE* mutants: influence on lipopolysaccharide structure, colonization, and biofilm formation. *Infect. Immun.* **69**, 435-445.

Nubel,U., Reissbrodt,R., Weller,A., Grunow,R., Porsch-Ozcurumez,M., Tomaso,H., Hofer,E., Splettstoesser,W., Finke,E.-J., Tschape,H. and Witte,W. (2006) Population structure of *Francisella tularensis*. *J. Bacteriol.* **188**, 5319-5324.

Nunoshiba,T., deRojas-Walker,T., Wishnok,J.S., Tannenbaum,S.R. and Demple,B. (1993) Activation by nitric oxide of an oxidative-stress response that defends *Escherichia coli* against activated macrophages. *Proc Natl Acad Sci U S A.* **90**, 9993-9997.

Oh,K.J., Park,Y.S., Lee,K.A., Chung,Y.J. and Cho,T.J. (2004) Molecular characterization of a *thiJ*-like gene in Chinese cabbage. *J. Biochem. Mol. Biol.* **37**, 343-350.

Ohara,Y., Sato,T. and Homma,M. (1974) Serological studies on *Francisella tularensis*, *Francisella novicida*, *Yersinia philomiragia*, and *Brucella abortus*. *Int. J. Syst. Bacteriol.* **24**, 191-196.

Ohara,Y., Sato,T., Fujita,H., Ueno,T. and Homma,M. (1991) Clinical manifestations of tularemia in Japan - analysis of 1355 cases observed between 1924 and 1996. *Infection* **19**, 14-17.

Olsufjev,N.G., Emelyanova,O.S. and Dunaeva,T.N. (1959) Comparative study of strains of *B. tularense* in the Old and New World and their taxonomy. *J. Hyg. Epidemiol. Microbiol. Immunol.* **3**, 138-149.

Ottem,K.F., Nylund,A., Karlsbakk,E., Friis-Møller,A., Krossøy,B. and Knappskog,D. (2007) New species in the genus *Francisella* (Gammaproteobacteria; *Francisellaceae*); *Francisella piscicida* sp. nov. isolated from cod (*Gadus morhua*). *Arch. Microbiol.* **188**, 547-550.

Outten,F.W., Djaman,O. and Storz,G. (2004) A *suf* operon requirement for Fe-S cluster assembly during iron starvation in *Escherichia coli*. *Mol. Microbiol.* **52**, 861-872.

Owen, C. R. (1974) Genus *Francisella*. In *Bergey's manual or determinative bacteriology* ed. Buchanan,R.E. and Gibbons,N.E. pp. 283-285. Baltimore: The Williams and Wilkins Co.

Pallen,M.J. and Wren,B.W. (2007) Bacterial pathogenomics. *Nature* **449**, 835-842.

Pang,P.P. and Walker,G.C. (1983) Identification of the *uvrD* gene product of *Salmonella typhimurium* LT2. *J. Bacteriol.* **153**, 1172-1179.

Parkhill,J., Wren,B.W., Mungall,K.L., Ketley,J.M., Churcher,C., Basham,D., Chillingworth,T., Davies,R.M., Feltwell,T., Holroyd,S., Jagels,K., Karlyshev,A.V., Moule,S., Pallen,M.J., Penn,C.W., Quail,M., Rajandream,M.A., Rutherford,K., van Vliet,A.H., Whitehead,S. and Barrell,B.G. (2000) The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature* **403**, 665-668.

Parkhill,J., Wren,B.W., Thomson,N.R., Titball,R.W., Holden,M.T., Prentice,M.B., Sebaihia,M., James,K.D., Churcher,C., Mungall,K.L., Baker,S., Basham,D., Bentley,S.D., Brooks,K., Cerdeno-Tárraga,A.M., Chillingworth,T., Cronin,A., Davies,R.M., Davis,P., Dougan,G., Feltwell,T., Hamlin,N., Holroyd,S., Jagels,K., Karlyshev,A.V., Leather,S., Moule,S., Oyston,P.C.F., Quail,M., Rutherford,K., Simmonds,M., Skelton,J., Stevens,K., Whitehead,S. and Barrell,B.G. (2001) Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* **413**, 523-527.

Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D., Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T., Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T., Connerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N., Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S., Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P., Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J., Stevens,K., Whitehead,S. and Barrell,B.G. (2001) Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature* **413**, 848-852.

Patzer,S.I. and Hantke,K. (1999) SufS is a NifS-like protein, and SufD is necessary for stability of the [2Fe-2S] FhuF protein in *Escherichia coli*. *J. Bacteriol.* **181**, 3307-3309.

Pechous,R., Celli,J., Penoske,R., Hayes,S.F., Frank,D.W. and Zahrt,T.C. (2006) Construction and characterization of an attenuated purine auxotroph in a *Francisella tularensis* live vaccine strain. *Infect. Immun.* **74**, 4452-4461.

Perego,M., Higgins,C.F., Pearce,S.R., Gallagher,M.P. and Hoch,J.A. (1991) The oligopeptide transport system of *Bacillus subtilis* plays a role in the initiation of sporulation. *Mol. Microbiol.* **5**, 173-185.

Perez,J.C. and Groisman,E.A. (2007) Acid pH activation of the PmrA/PmrB two-component regulatory system of *Salmonella enterica*. *Mol. Microbiol.* **63**, 283-293.

Perrin,A., Nassif,X. and Tinsley,C. (1999) Identification of regions of the chromosome of *Neisseria meningitidis* and *Neisseria gonorrhoeae* which are specific to the pathogenic *Neisseria* species. *Infect. Immun.* **67**, 6119-6129.

Perrotte-Piquemal,M., Danchin,A. and Biville,F. (1999) Pyrophosphate increases the efficiency of enterobactin-dependent iron uptake in *Escherichia coli*. *Biochimie.* **81**, 245-253.

Pianzzola,M.J., Soubes,M. and Touati,D. (1996) Overproduction of the *rbo* gene product from *Desulfovibrio* species suppresses all deleterious effects of lack of superoxide dismutase in *Escherichia coli*. *J. Bacteriol.* **178**, 6736-6742.

Poole,K., Krebes,K., McNally,C. and Neshat,S. (1993) Multiple antibiotic resistance in *Pseudomonas aeruginosa*: evidence for involvement of an efflux operon. *J. Bacteriol.* **175**, 7363-7372.

Pratt,L.A. and Silhavy,T.J. (1995) Identification of base pairs important for OmpR-DNA interaction. *Mol. Microbiol.* **17**, 565-573.

Pusch,W., Wurmbach,J.H., Thiele,H. and Kostrzewa,M. (2002) MALDI-TOF mass spectrometry-based SNP genotyping. *Pharmacogenomics* **3**, 537-548.

Rachman,H., Strong,M., Ulrichs,T., Grode,L., Schuchhardt,J., Mollenkopf,H., Kosmiadi,G.A., Eisenberg,D. and Kaufmann,S.H.E. (2006) Unique transcriptome signature of *Mycobacterium tuberculosis* in pulmonary tuberculosis. *Infect. Immun.* **74**, 1233-1242.

Ramagopal,S. and Subramanian,A.R. (1974) Alteration in the acetylation level of ribosomal protein L12 during growth cycle of *Escherichia coli. Proc Natl Acad Sci U S A.* **71**, 2136-2140.

Rebeil,R., Sun,Y., Chooback,L., Pedraza-Reyes,M., Kinsland,C., Begley,T.P. and Nicholson,W.L. (1998) Spore photoproduct lyase from *Bacillus subtilis* spores is a novel iron-sulfur DNA repair enzyme which shares features with proteins such as class III anaerobic ribonucleotide reductases and pyruvate-formate lyases. *J. Bacteriol.* **180**, 4879-4885.

Reid,J.D. and Hunter,C.N. (2002) Current understanding of the function of magnesium chelatase. *Biochem. Soc. Trans.* **30**, 643-645.

Reilly,T.J., Baron,G.S., Nano,F.E. and Kuhlenschmidt,M.S. (1996) Characterization and sequencing of a respiratory burst-inhibiting acid phosphatase from *Francisella tularensis. J. Biol. Chem.* **271**, 10973-10983.

Reintjes,R., Dedushaj,I., Gjini,A., Jorgensen,T.R., Cotter,B., Lieftucht,A., D'Ancona,F., Dennis,D.T., Kosoy,M.A., Mulliqi-Osmani,G., Grunow,R., Kalaveshi,A., Gashi,L. and Humolli,I. (2002) Tularemia outbreak investigation in Kosovo: case control and environmental studies. *Emerg. Infect. Dis.* **8**, 69-73.

Ricci,S., Janulczyk,R. and Bjorck,L. (2002) The regulator PerR is involved in oxidative stress response and iron homeostasis and is necessary for full virulence of *Streptococcus pyogenes*. *Infect. Immun.* **70**, 4968-4976.

Ritter,D.B. and Gerloff,R.K. (1966) Deoxyribonucleic acid hybridization among some species of the genus *Pasturella*. *J. Bacteriol.* **92**, 1838-1839.

Robertson,G.T. and Roop Jr.,R.M. (1999) The *Brucella abortus* host factor I (HF-I) protein contributes to stress resistance during stationary phase and is a major determinant of virulence in mice. *Mol. Microbiol.* **34**, 690-700.

Rocha,E.R., Selby,T., Coleman,J.P. and Smith,C.J. (1996) Oxidative stress response in an anaerobe, *Bacteroides fragilis*: a role for catalase in protection against hydrogen peroxide. *J. Bacteriol.* **178**, 6895-6903.

Rogutskii,S.V., Khramtsov,M.M., Avchinikov,A.V., Gar'kavaia,O.I., Larchenkova,N.V., Golovanov,A.M., Parfenova,T.A. and Rudkovskaia,L.B. (1997) An epidemiological investigation of a tularemia outbreak in Smolensk Province. *Zh. Mikrobiol. Epidemiol. Immunobiol.* **2**, 33-37.

Rohmer,L., Brittnacher,M., Svensson,K., Buckley,D., Haugen,E., Zhou,Y., Chang,J., Levy,R., Hayden,H., Forsman,M., Olson,M., Johansson,A., Kaul,R. and Miller,S.I. (2006) Potential source of *Francisella tularensis* live vaccine strain attenuation determined by genome comparison. *Infect. Immun.* **74**, 6895-6906.

Rosen,H. and Klebanoff,S.J. (1985) Oxidation of microbial iron-sulfur centers by the myeloperoxidase-H_2O_2-halide antimicrobial system. *Infect. Immun.* **47**, 613-618.

Rouquette,C., de Chastellier,C., Nair,S. and Berche,P. (1998) The ClipC ATPase of *Listeria monocytogenes* is a general stress protein required for virulence and promoting early bacterial escape from the phagosome of macrophages. *Mol. Microbiol.* **27**, 1235-1245.

Rusnak,J.M., Kortepeter,M.G., Hawley,R.J., Anderson,A.O., Boudreau,E. and Eitzen,E. (2004) Risk of occupationally acquired illnesses from biological threat agents in unvaccinated laboratory workers. *Biosecur. Bioterror.* **2**, 281-293.

Saier Jr.,M.H. (1977) Bacterial phosphoenolpyruvate: sugar phosphotransferase systems: structural, functional, and evolutionary interrelationships. *Bacteriol. Rev.* **41**, 856-871.

Samrakandi,M.M., Zhang,C., Nietfeldt,J., Kim,J., Iwen,P.C., Olsen,M.E., Fey,P.D., Duhamel,G.E., Hinrichs,S.H., Cirillo,J.D. and A.K.Benson (2004) Genome diversity among regional populations of *Francisella tularensis* subspecies *tularensis* and *Francisella tularensis* subspecies *holarctica* isolated from the US. *FEMS Microbiol. Lett.* **237**, 9-17.

Sandström,G., Löfgren,S. and Tärnvik,A. (1988) A capsule-deficient mutant of *Francisella tularensis* LVS exhibits enhanced sensitivity to killing by serum but diminished sensitivity
to kKilling by polymorphonuclear leukocytes. *Infect. Immun.* **56**, 1194-1202.

Sandström,G., Sjöstedt,A., Forsman,M., Pavlovich,N.V. and Mishankin,B.N. (1992) Characterization and classification of strains of *Francisella tularensis* isolated in the Central Asian focus of the Soviet Union and in Japan. *J. Clin. Microbiol.* **30**, 172-175.

Santos,P.M., Benndorf,D. and Sá-Correia,I. (2007) Insights into *Pseudomonas putida* KT2440 response to phenol-induced stress by quantitative proteomics. *Proteomics* **4**, 2640-2652.

Sarkar,N. (1997) Polyadenylation of mRNA in bacteria. *Microbiol.* **142**, 3125-3133.

Sarria,J.C., Vidal,A.M., Kimbrough,R.C. and Figueroa,J.E. (2003) Fatal infection caused by *Francisella tularensis* in a neutropenic bone marrow transplant recipient. *Ann Hematol.* **82**, 42-43.

Saslaw,S., Eigelsbach,H.T., Wilson,H.E., Prior,J.A. and Carhart,S. (1961) Tularemia vaccine study. I. Intracutaneous challenge. *Arch. Intern. Med.* **107**, 689-701.

Saslaw,S., Eigelsbach,H.T., Prior,J.A., Wilson,H.E. and Carhart,S. (1961) Tularemia vaccine study. II. Respiratory challenge. *Arch. Intern. Med.* **107**, 702-714.

Schoolnik,G.K. (2002) Functional and comparative genomics of pathogenic bacteria. *Curr. Opin. Microbiol.* **5**, 20-26.

Serres,M.H., Goswami,S. and Riley,M. (2004) GenProtEC: an updated and improved analysis of functions of *Escherichia coli* K-12 proteins. *Nuc. Acid. Res.* **32**, D300-D302.

Shalon,D., Smith,S.J. and Brown,P.O. (1996) A DNA system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res.* **6**, 639-645.

Sjöstedt,A., Conlan,J.W. and North,R.J. (1994) Neutrophils are critical for host defense against primary infection with the facultative intracellular bacterium *Francisella tularensis* in mice and participate in defense against reinfection. *Infect. Immun.* **62**, 2779-2783.

Sjöstedt,A. (2006) Intracellular survival mechanisms of *Francisella tularensis*, a stealth pathogen. *Microbes Infect.* **8**, 561-567.

Soncini,F.C. and Groisman,E.A. (1996) Two-component regulatory systems can interact to process multiple environmental signals. *J. Bacteriol.* **178**, 6796-6801.

Soule,M., Cain,K., LeFrentz,S. and Call,D.R. (2005) Combining suppression subtractive hybridization and microarrays to map the intraspecies phylogeny of *Flavobacterium psychrophilum*. *Infect. Immun.* **73**, 3799-3802.

Splettstoesser,W.D., Matz-Rensing,K., Seibold,E., Tomaso,H., Al Dahouk,S., Grunow,R., Essbauer,S., Buckendahl,A., Finke,E.J. and Neubauer,H. (2007) Re-emergence of *Francisella tularensis* in Germany: fatal tularaemia in a colony of semi-free-living marmosets (*Callithrix jacchus*). *Epidemiol. Infect.* **135**, 1256-1265.

Squires,C.L., Pedersen,S., Ross,B.M. and Squires,C. (1991) ClpB is the *Escherichia coli* heat shock protein F84.1. *J. Bacteriol.* **173**, 4254-4262.

Stabler,R.A., Hinds,J., Witney,A.A., Isherwood,K.E., Oyston,P.C.F., Titball,R.W., Wren,B.W., Hinchliffe,S., Prentice,M.B., Mangan,J.A. and Butcher,P.D. (2003) Construction of a *Yersinia pestis* microarray. *Adv. Exp. Med. Biol.* **529**, 49.

Stadtman,E.R. (1990) Metal ion-catalyzed oxidation of proteins: biochemical mechanism and biological consequences. *Free Rad. Biol.* **9**, 315-325.

Stef,M., Simon,D., Burgelin,I., Guisle,I., Chevalier,C., Delrue,M.A., Lacombe,D., Léger,J. and Arveiler,B. (2006) Testing and improving experimental parameters for the use of low molecular weight targets in array-CGH experiments. *Hum. Mutat.* **27**, 1143-1150.

Steinemann,T.L., Sheikholeslami,M.R., Brown,H.H. and Bradsher,R.W. (1999) Oculoglandular tularemia. *Arch. Opthalmol.* **117**, 132-133.

Stewart,G.R., Wernisch,L., Stabler,R., Mangan,J.A., Hinds,J., Laing,K.G., Young,D.B. and Butcher,P.D. (2002) Dissection of the heat-shock response in *Mycobacterium tuberculosis* using mutants and microarrays. *Microbiol.* **148**, 3129-3138.

Stohl,E.A., Criss,A.K. and Seifert,S.H. (2005) The transcriptome response of *Neisseria gonorrhoeae* to hydrogen peroxide reveals genes with previously uncharacterized roles in oxidative damage protection. *Mol. Microbiol.* **58**, 520-532.

Su,J., Yang,J., Zhao,D., Kawula,T.H., Banas,J.A. and Zhang,J.R. (2007) Genome-wide identification of *Francisella tularensis* virulence determinants. *Infect. Immun.* **75**, 3089-3101.

Sullivan,J.T., Jeffery,E.F., Shannon,J.D. and Ramakrishnan,G. (2006) Characterization of the siderophore of *Francisella tularensis* and the role of *fslA* in siderophore production. *J. Bacteriol.* **188**, 3785-3795.

Sunderrajan,E.V., Hutton,J. and Marienfeld,R.D. (1985) Adult respiratory distress syndrome secondary to tularemia pneumonia. *Arch. Intern. Med.* **145**, 1435-1437.

Suzuki,T., Murai,T., Fukuda,I., Tobe,T., Yoshikawa,M. and Sasakawa,C. (1994) Identification and characterization of a chromosomal virulence gene *vacJ* required for intercellular spreading of *Shigella flexneri*. *Mol. Microbiol.* **11**, 31-41.

Svensson,K., Larsson,P., Johansson,D., Bystrom,M., Forsman,M. and Johansson,A. (2005) Evolution of subspecies of *Francisella tularensis*. *J. Bacteriol.* **187**, 3903-3908.

Talaat,A.M., Hunter,P. and Johnston,S.A. (2000) Genome-directed primers for selective labeling of bacterial transcripts for DNA microarray analysis. *Nature Biotechnology* **18**, 679-682.

Taljanidisz,J., Shen,P. and Sarkar,N. (1997) Half-life of *Escherichia coli* polyadenylated lipoprotein mRNA. *Biochem. Mol. Biol. Int.* **42**, 211-215.

Tao,X. and Tong,L. (2003) Crystal structure of human DJ-1, a protein associated with early onset Parkinson's disease. *J. Biol. Chem.* **278**, 31372-31379.

Tärnvik,A. (1989) Nature of protective immunity to *Francisella tularensis*. *Rev. Infect.Dis.* **11**, 440-447.

Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C., Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F., Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D., Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D., Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E., Cittone,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H., Qin,H., Vamahevan,J., Gill,J., Scarlato,V., Masignani,V., Pizza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R., Rappuoli,R. and Venter,J.C. (2000) Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. *Science* **287**, 1809-1815.

Thomas,R.M., Titball,R.W., Oyston,P.C.F., Griffin,K.F., Waters,E., Hitchen,P.G., Michell,S.L., Grice,I.D., Wilson,J.C. and Prior,J.L. (2007) The immunologically distinct O-antigens from *Francisella tularensis* subspecies *tularensis* and *Francisella novicida* are both virulence determinants and protective antigens. *Infect. Immun.* **75**, 371-378.

Tigertt,W.D. (1962) Soviet viable *Pasturella tularensis* vaccines: a review of selected articles. *Bacteriol. Rev.* **26**, 354-373.

Tuanyok,A., Stanley,K.H., Nierman,W., Yu,Y., Dunbar,J., Moore,R.A., Baker,P., Tom,M., Ling,J.M.L. and Woods,D.E. (2005) Genome-wide expression analysis of iron regulation in *Burkholderia pseudomallei* and *Burkholderia mallei* using DNA microarrays. *FEMS Microbiol. Lett.* **252**, 327-335.

van Vliet,A.H., Baillon,M.A., Penn,C.W. and Ketley,J.M. (2001) The iron-induced ferredoxin FdxA of *Campylobacter jejuni* is involved in aerotolerance. *FEMS Microbiol. Lett.* **196**, 189-193.

Vasil,M.L. (2007) How we learnt about iron acquisition in *Pseudomonas aeruginosa*: a series of very fortunate events. *Biometals* **20**, 587-601.

Villarroya,M., Pérez-Roger,I., Macián,F. and Armegod,M.E. (1998) Stationary phase induction of *dnaN* and *recF*, two genes of *Escherichia coli* involved in DNA replication and repair. *EMBO J.* **17**, 1829-1837.

Volkman,S.K., Hartl,D.L., Wirth,D.F., Nielsen,K.M., Choi,M., Batalov,S., Zhou,Y., Plouffe,D., Le Roche,K.G., Abagyan,R. and Winzeler,E.A. (2002) Excess polymorphisms in genes for membrane proteins in *Plasmodium falciparum*. *Science* **298**, 216-218.

Wang,G. and Maier,R.J. (2004) An NADPH quinone reductase of *Helicobacter pylori* plays an important role in oxidative stress resistance and host colonization. *Infect. Immun.* **72**, 1391-1396.

Wang,W. and Bechhofer,D.H. (1997) *Bacillus subtilis* RNase III gene: cloning, function of the gene in *Escherichia coli*, and construction of *Bacillus subtilis* strains with altered *rnc* loci. *J. Bacteriol.* **179**, 7379-7385.

Ward,P.P., Mendoza-Meneses,M., Cunningham,G.A. and Conneely,O.M. (2003) Iron status in mice carrying a targeted disruption of lactoferrin. *Mol. Cell. Biol.* **23**, 178-185.

Weiss,D.S., Brotcke,A., Henry,T., Margolis,J.J., Chan,K. and Monack,D.M. (2007) *In vivo* negative selection screen identifies genes required for *Francisella* virulence. *Proc Natl Acad Sci U S A.* **104**, 6037-6042.

Weissborn,A.C., Rumley,M.K. and Kennedy,E.P. (1992) Isolation and characterization of *Escherichia coli* mutants blocked in production of membrane-derived oligosaccharides. *J Bact* **174**, 4856-4859.

Weissborn,A.C., Liu,Q., Rumley,M.K. and Kennedy,E.P. (1994) UTP: alpha-D-glucose-1-phosphate uridylyltransferase of *Escherichia coli*: isolation and DNA sequence of the *galU* gene and purification of the enzyme. *J Bact* **176**, 2611-2618.

Wenger JD. (1989) Infection caused by *Francisella philomiragia* (formerly *Yersinia philomiragia*). A newly recognized human pathogen. *Ann. Internat. Med.* **110**, 888-892.

Whipp,M.J., Davis,J.M., Lum,G., Boer,J., Zhou,Y., Bearden,S.W., Petersen,J.M., Chu,M.C. and Hogg,G. (2003) Characterization of a *novicida*-like subspecies of *Francisella tularensis* isolated in Australia. *J. Microbiol. Methods* **52**, 839-842.

Wik,O. (2006) Large tularemia outbreak in Varmland, central Sweden, 2006. *Euro Surveill.* **11**, E060921.1.

Wilson, G. S. and Miles, A. A. (1964) *Brucella tularensis.* In *Principles of Bacteriology and Immunity* ed. Topley and Wilson pp. 1013-1014. Baltimore: The Williams and Wilkins Co.

Wilson,J.W., Ramamurthy,R., Porwollik,S., McClelland,M., Hammond,T., Allen,P., Ott,C.M., Pierson,D.L. and Nickerson,C.A. (2002) Microarray analysis identified *Salmonella* genes belonging to the low-shear modelled microgravity regulon. *Proc Natl Acad Sci U S A.* **99**, 13807-13812.

Winzeler,E.A., Castillo-Davis,C.I., Oshiro,G., Liang,D., Richards,D.R., Zhou,Y. and Hartl,D.L. (2003) Genetic diversity in yeast assessed with whole-genome oligonucleotide arrays. *Genetics* **163**, 79-89.

Wolfson,J.S., Hooper,D.C., McHugh,G.L., Bozza,M.A. and Swartz,M.N. (1990) Mutants of *Escherichia coli* K-12 exhibiting reduced killing by both quinolone and beta-lactam antimicrobial agents. *Antimicrob. Agents Chemo.* **34**, 1938-1943.

Wong,C.W., Albert,T.J., Vega,V.B., Norton,J.E., Cutler,D.J., Richmond,T.A., Stanton,L.W., Liu,E.T. and Miller,L.D. (2004) Tracking the evolution of the SARS coronavirus using high-throughput, high-density resequencing arrays. *Genome Res.* **14**, 398-405.

Yamamoto,M., Nomura,M., Ohsawa,H. and Maruo,B. (1977) Identification of a temperature-senstive asparaginyl-transfer ribonucleic acid synthetase mutant of *Escherichia coli. J. Bacteriol.* **132**, 127-131.

Yang,Y.H., ,D.S., Luu,P., Lin,D.M., Peng,V., Ngai,J. and Speed,T.P. (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nuc. Acid. Res.* **30**, 1-11.

Yanofsky,C., Horn,V. and Nakamura,Y. (1996) Loss of overproduction of polypeptide release factor 3 influences expression of the tryptophanase operon of *Escherichia coli*. *J. Bacteriol.* **178**, 3755-3762.

Yoshikawa,A., Isono,S., Sheback,A. and Isono,K. (1987) Cloning and nucleotide sequencing of the genes *riml* and *rimJ* which encode enzymes acetylating ribosomal proteins S18 and S5 of *Escherichia coli* K12. *Mol Gen Genet* **209**, 481-8.

Zaidi,S.A. and Singer,C. (2002) Gastrointestinal and hepatic manifestations of tick-borne diseases in the United States. *Clin. Infect. Dis.* **34**, 1206-1212.

Zavilgelsky,G.B., Kotova,V.Y., Mazhul',M.M. and Manukhov,I.V. (2002) Role of HSP70 (DnaK-DnaJ-GrpE) and HSP100 (ClpA and ClpB) chaperones in refolding and increased thermal stability of bacterial luciferases in *Escherichia coli* cells. *Biochem. (Mosc.)* **67**, 986-992.

Zhang,F., Liu,W., Chu,M.C., He,J., Duan,Q., Wu,X.M., Zhang,P.H., Zhao,Q.M., Xin,Z.T. and Cao,W.C. (2006) *Francisella tularensis* in rodents, China. *Emerg. Infect. Dis.* **12**, 994-996.

Zhang,Z., Gosset,G., Barabote,R., Gonzalez,C.S., Cuevas,W.A. and Saier Jr.,M.H. (2005) Functional interactions between the carbon and iron utilization regulators Crp and Fur, in *Escherichia coli*. *J Bacteriol.* **187**, 980-990.

Zheng,L., Cash,V.L., Flint,D.H. and Dean,D.R. (1998) Assembly of iron-sulfur clusters. Identification of an *iscSUA-hscBA-fdx* gene cluster from *Azotobacter vinelandii*. *J. Biol. Chem.* **273**, 13264-13272.

Zwick,M.E., Mcafee,F., Cutler,D.J., Read,T.D., Ravel,J., Bowman,G.R., Galloway,D.R. and Mateczun,A. (2005) Microarray-based resequencing of multiple *Bacillus anthracis* isolates. *Genome Biol.* **6**, R10.